

参赛密码 _____
(由组委会填写)

第十一届华为杯全国研究生数学建模竞赛

学 校	同济大学
参赛队号	10247034
队员姓名	1.高心健
	2.周辰红
	3.刘园珍

参赛密码 _____
(由组委会填写)



第十一届华为杯全国研究生数学建模竞赛

题 目 人体营养健康角度的中国果蔬发展战略研究

摘 要：

本文针对中国果蔬发展战略的建模问题，从居民营养健康角度出发，运用真实数据主要完成了以下几方面的工作：

对于问题一，本文首先对查找到的数据进行预处理，并对数据进行相关性分析和粗大误差分析。然后，针对果蔬品种的营养素成分和含量，建立分级打分制度，并定义了“营养价值”、“常见度”和“研究价值”等相关指标，筛选出主要的 8 种水果和 6 种蔬菜作为研究对象。其次，运用回归模型预测各果蔬品种的年产量，进而预估消费量。最后，为了使预测结果更加准确，利用 Logistic 模型对回归模型预测结果进行修正，从而改进了回归模型的产量预测结果。

对于问题二，本文首先通过查阅文献资料，整理营养素的摄入大体现状；然后根据附件中营养素参考日摄入量，按照加权平均法计算得到人均日摄入量标准；其次，利用问题一中 Logistic 模型预测出的果蔬品种消费量，计算目前中国居民人均营养素摄入水平，并与摄入量标准对比，引入营养素“达标指数”，评价我国居民营养素摄入是否合理；最后，根据各营养素“达标指数”趋势，预测至 2020 年我国居民的人体营养健康状况。

问题三是一个线性规划问题，求解在保证营养均衡的条件下，主要果蔬品种的按年度合理人均消费量，使居民的购买成本最低。首先，本文将全国分为南北两个区域，并分春、夏、秋、冬四季，引入应地系数和应季系数，研究地区和季节对果蔬价格的影响。然后，利用 Matlab 求解该线性规划问题，给出南北方四季按年度合理的果蔬品种人均消费量。最后，利用聚类分析方法，对不同的果蔬品种按照营养素种类进行聚类，充分考虑果蔬品种之间的相似性和互补性，并给出按照价格从小到大的果蔬替代消费方案。

问题四是一个多目标规划问题，需要在综合考虑居民人体营养均衡、土地

面积等因素的情况下，一方面使得居民购买果蔬成本达到最低，另一方面使得种植者的利润最大。首先，本文通过线性加权法对 2015-2020 年的果蔬产品的年度合理人均消费量进行求解；然后，为了获得国家对于果蔬品种生产规模的战略调整方案，不断提高我国居民的营养素摄入水平，将各项营养素的“达标指数”作为动态约束条件引入优化模型中；最后，通过每年不断调整营养素达标指数约束，使得居民各项营养素摄入量可以逐步均衡，实现全部营养素摄入水平更快达标的目的，为国家的宏观调控提供理论依据。

关键字：相关性分析 线性规划 粗大误差分析 回归模型 logistic 模型 聚类分析 达标指数 多目标规划

目录

1. 问题重述.....	5
1.1 问题背景.....	5
1.2 待解决的问题.....	5
2. 问题分析.....	6
3. 符号说明及模型假设.....	7
3.1 符号说明.....	7
3.2 模型基本假设.....	8
4. 主要果蔬的消费量建模.....	8
4.1 确定主要水果.....	10
4.1.1 确定研究的营养素范围	10
4.1.2 数据预处理	10
4.1.3 水果营养价值评分标准制定	11
4.1.4 水果研究价值标准制定	12
4.1.5 确定主要水果	13
4.2 水果消费量建模.....	14
4.2.1 基于相关分析的变量分析	14
4.2.2 水果年消费量估计及发展趋势	14
4.3 确定主要蔬菜.....	23
4.3.1 数据预处理	23
4.3.2 粗大误差分析	25
4.3.3 运用分段评分法确定主要蔬菜	30
4.4 估计主要蔬菜消费量.....	31
4.4.1 多项式拟合法	32
4.4.2 Logistic 复合模型法	33
5 营养素摄入现状与摄入水平探究.....	34
5.1 营养素摄入现状.....	34
5.1.1 营养素摄入现状文献分析	34
5.1.2 营养素摄入量建模分析	35
5.2 我国居民营养素摄入水平分析.....	36
5.2.1 我国居民营养素摄入量计算	36
5.2.2 达标指数确定	37
5.2.3 我国居民摄入营养素水平分析	37
5.3 2020 年我国居民营养健康状况.....	38
6 居民果蔬人均消费优化方案.....	39
6.1 居民果蔬人均消费优化准备.....	40
6.1.1 分区域分季节	40
6.1.2 应地系数和应季系数的引入	40
6.2 居民果蔬人均消费优化求解.....	41
6.3 对水果和蔬菜进行聚类分析.....	42

6.3.1 欧氏距离	42
6.3.2 马氏距离	43
6.3.3 运用 SPSS 对果蔬聚类	43
6.4 我国居民人均果蔬消费搭配与替换方案	44
7 问题四模型建立与求解	45
7.1 多目标规划模型的建立	45
7.2 动态线性加权法及其实现	46
7.3 动态多目标规划模型的求解	47
7.4 国家宏观调控战略	49
8 政策建议	49
8.1 政策建议	49
9 模型评价	50
9.1 模型的优点	50
9.2 模型的缺点	51
参考文献	52

1. 问题重述

1.1 问题背景

近年来，我国农产品综合生产能力稳步提高，食物供需基本平衡，食品安全状况总体稳定向好，居民营养健康状况明显改善，食物与营养发展成效显著。但是，我国食物生产还不能适应营养需求，居民营养不足与过剩并存，营养与健康知识缺乏。为规划解决这一问题，2014 年 1 月 28 日国务院办公厅正式颁布《中国食物与营养发展纲要（2014—2020 年）》。纲要指出，我国要推广膳食结构多样化的健康消费模式，保障充足的能量和蛋白质摄入量，控制脂肪摄入量，保持适量的维生素和矿物质摄入量，充分保证人体营养素的摄入量。

人体需要的营养素主要有蛋白质、脂肪、维生素、矿物质、糖和水。其中维生素对于维持人体新陈代谢的生理功能是不可或缺的，多达 30 余种，分为脂溶性维生素（如维生素 A、D、E、K 等）和水溶性维生素（如维生素 B1、B2、B6、B12、C 等）。矿物质无机盐等亦是构成人体的重要成分，约占人体体重的 5%，主要有钙、钾、硫等以及微量元素铁、锌等。另外适量地补充膳食纤维对促进良好的消化和排泄固体废物有着举足轻重的作用。

水果和蔬菜都是富含营养素的最主要的农产品，主要为人体提供矿物质、维生素、膳食纤维。近年来，我国水果和蔬菜种植面积和产量迅速增长，水果和蔬菜品种也日益丰富。但是伴随着我国居民的喜食、饱食、偏食的饮食习惯，果蔬消费（品种和数量）就显得盲目无序，进而影响到果蔬的生产。

1.2 待解决的问题

● 问题一

科学决策的基础是比较准确地掌握情况。但我国蔬菜和水果品种繁多，无论是中国官方公布的数据还是世界粮农组织（FAO）、美国农业部（USDA）等发布的数据均不完整，缺失较为普遍，而且品种、口径不一。我们既不可能也没有必要了解全部数据，对这样的宏观问题，恰当的方法是选取主要的水果和蔬菜品种进行研究。因此，要求主要的水果、蔬菜品种不仅总计产量应分别超过它们各自总产量的 90%，而且这部分品种所蕴含营养素无论在成分上还是在含量上都满足研究的需要。运用数学手段从附件表格中筛选出主要的水果和蔬菜品种，并尝试用多种方法建立数学模型对其消费量进行估计，研究其发展趋势。

● 问题二

摸清我国居民矿物质、维生素、膳食纤维等营养素摄入现状。请结合为保障人体健康所需要的各种营养成分的范围（见附件和参考文献）和前面预测的人均消费结果，评价中国居民目前矿物质、维生素、膳食纤维等营养的年摄入水平是否合理。按照水果和蔬菜近期的消费趋势，至 2020 年，中国居民的人体营养健康状况是趋于好转还是恶化？请给出支持结论的充分依据。

● 问题三

不同的蔬菜、水果尽管各种营养素含量各不相同，但营养素的种类大致相

近，存在着食用功能的相似性。所以，水果与水果之间、蔬菜与蔬菜之间、水果与蔬菜之间从营养学角度在一定程度上可以相互替代、相互补充。由于每种蔬菜、水果所含有的维生素、矿物质、膳食纤维成分、含量不尽相同，价格也有差异，因而在保证营养均衡满足健康需要条件下，如何选择消费产品是个普遍的问题。请你们为当今中国居民（可以分区域分季节）提供主要的水果和蔬菜产品的按年度合理人均消费量，使人们能够以较低的购买成本（假定各品种价格按照原有趋势合理变动）满足自身的营养健康需要。

● 问题四

为实现人体营养均衡满足健康需要，国家可能需要对水果和蔬菜各品种的生产规模做出战略性调整。一方面国家要考虑到居民人体的营养均衡，并使营养摄入量尽量在合理范围内；另一方面也要顾及居民的购买成本，使其购买成本尽量低；同时还要使种植者能够尽量获得较大收益；而且，作为国家宏观战略，还要考虑进出口贸易、土地面积等其他因素。请你们基于上述考虑，建立数学模型重新计算中国居民主要的水果和蔬菜产品的按年度合理人均消费量，并给出到 2020 年我国水果和蔬菜产品生产的调整战略。

● 问题五

结合前面的研究结论，给相关部门提供 1000 字左右的政策建议。

2. 问题分析

蔬菜和水果富含维生素等各种营养素，然而我国居民的饮食习惯偏向于挑食，喜食，厌食，这就可能造成营养素不均衡。因此，研究蔬菜水果的消费量及对消费情况进行建模分析是非常有必要的。根据题中所给背景及问题材料，我们就五个问题首先进行如下分析：

● 问题一

问题一要求合乎现状，建立模型对水果、蔬菜的消费量进行估计。我国蔬菜和水果品种繁多，相应统计机关及口径多种多样，不仅有中国农业部统计的官方数据，也有世界粮农组织、美国农业部等相关部门发布的数据资料。对于多种多样的统计口径，首先我们选定中华人民共和国农业部种植业管理司给出的数据，然后根据现有的文献的研究成果及检索到的相关年鉴数据，从实际出发，运用现有数据，创新定义“营养价值”和“研究价值”并结合新定义的“常见度”指标对水果和蔬菜打分，从而选取主要水果 8 种、主要蔬菜 6 种。题中附件给出了水果或蔬菜的损耗率，因而可以根据预测产量预估消费量。为此，我们从现实出发，运用回归模型分别预测年产量：线性回归，多项式回归；并且利用 Logistic 模型修正改进了回归模型给出的产量预测结果。三种模型方法具有不同的特点，都有较高的有效性与精确性。总结来说，本题建模方法主要包括：线性回归，多项式回归，Logistic 复合模型。

● 问题二

问题二要求摸清我国居民矿物质、维生素、膳食纤维的摄入现状，以及评价我国居民目前营养素的年摄入水平是否合理，最后估计 2020 年，我国居民的营养健康状况是趋于好转还是恶化，并要求给出支持结论的充分依据。针对问题要求，本文首先通过查阅文献资料，整理营养素的摄入大体现状，宏观上了解相关学者的研究结论；然后根据问题一得出的消费量进而估计预测人均摄入

水平，并且与标准的人均摄入量对比，引入营养素“达标指数”，从而可以评价我国居民营养素摄入是否合理；最后可以根据消费量趋势，评价我国居民的人体营养健康状况是趋于好转还是恶化。

● 问题三

问题三要求在保证营养均衡满足健康需要条件下，提供主要的水果和蔬菜产品的按年度合理人均消费量，使人们能够以较低的购买成本，这是一个线性规划问题。首先，我们将全国分为南北两个区域，并按照春、夏、秋、冬四季来研究该问题。为综合考虑这个问题，我们引入应地系数和应季系数来衡量地区和季节因素对价格的影响。然后，利用 Matlab 求解该优化模型的解，给出南北方四季的主要水果和蔬菜的按年度合理的人均消费量。最后，我们利用聚类分析方法，对不同的水果和蔬菜按照营养素种类进行聚类，进而对优化模型得到的果蔬方案进行各种组合搭配，并给出按照价格从小到大的替代方案。这样，居民可以按照各种喜好搭配自己的饮食方案，又可以选择到价格较低的搭配方案。

● 问题四

问题四是一个多目标规划问题，需要在综合考虑居民人体营养均衡的情况下，一方面使得居民购买成本达到最低，另一方面使得种植者的利润最大，可以通过线性加权法求解。为了获得国家对于水果和蔬菜各品种的生产规模的战略调整方案，对 2015-2020 年的果蔬产品的年度合理人均消费量进行求解，考虑营养均衡、土地面积等因素的影响，可以认为在国家对果蔬各品种调控前后的年总消费量不变，仅仅改变了各种果蔬品种的比例。为了达到居民人体营养均衡的目的，可以将各项营养素的达标指数作为动态约束条件，使得居民各项营养素摄入量可以逐步均衡，最终实现全部达标的目的。

● 问题五

结合题目前几问的研究结论，并通过阅读相关文献以及现有的一些政策法规，总结得出该给出的政策建议。

3. 符号说明及模型假设

3.1 符号说明

为了便于描述问题，我们用一些符号来替代问题中涉及的一些变量，如表 4-1 所示，其他一些变量在文中陆续说明。

表 1 符号含义表

符号	含义
C_{\min}	某种营养素在各种水果/蔬菜中含量的最小值
C_{mid}	某种营养素在各种水果/蔬菜中含量的中位数
C_{\max}	某种营养素在各种水果/蔬菜中含量的最大值
Com	常见度系数
y	具体的某一年份，数值即为年份值

Pro	年产量
$NVal$	营养价值
$RVal$	研究价值
ρ	相关系数
$w1$	田间地头到大市场损耗率
$w2$	大市场到零售市场损耗率
$w3$	零售市场到餐桌的损耗率
Con	水果每年的消费量
β	达标指数
ω	加权因子
λ	调整系数
a	营养素含量
b	标准摄入量
Con	水果每年的消费量

3.2 模型基本假设

对于所研究的问题的参数和范围，本文作出如下几点假设和限定：官方统计数据真实有效

- 通过网络在中华人民共和国农业部种植业管理司查询的相关数据真实有效，并能反映指标所代表的真实含义；
- 水果及蔬菜大类采用官方给出的类别，其中水果 12 种、蔬菜 9 种，不考虑其他种类的水果蔬菜；
- 每种蔬菜可食部分和不可食部分所含营养素含量相同；
- 我国人口没有大幅度改变，人口基数为 13 亿；
- 人体摄入的营养素全部被吸收；
- 水果的进口总量和出口总量相等；
- 蔬菜的进口总量和出口总量相等；
- 国家的宏观调控不影响每年果蔬总产量。

4. 主要果蔬的消费量建模

问题一要求建立模型对蔬菜水果的消费量进行估计并对其发展趋势进行研究。针对该问题，已有大量文献进行了研究。文献[1]对中国水果产量利用灰色马尔科夫模型进行预测；文献[2]在研究国外水果消费与城市化水平相关的基础上，着重探讨了我国水果消费与城市化水平的相关性，最后提出加快城市化、促进我国水果消费的策略；文献[3]从中国水果消费现状出发，分析了水果消费的特点和影响水果消费的因素，对中国居民中长期水果消费进行了预测。研究结果显示中国居民水果消费目前主要以鲜食为主，消费总量稳中上升；随着收入水平的提高和生活方式的变化，中国水果消费在总量上升的同时，进一步向多类型、多层次、多样化的质量安全型消费结构发展。

本问题本小文从消费现实出发，运用两大模型分别预测年产量：线性回归，多项式回归；并且创新地利用并改进 Logistic 模型修正改进了回归模型给出的产量预测结果。三种模型方法具有不同的特点，都有较高的有效性与精确性。

首先给出本小问的水果、蔬菜采用的技术路线图，如图 1、图 2 所示。

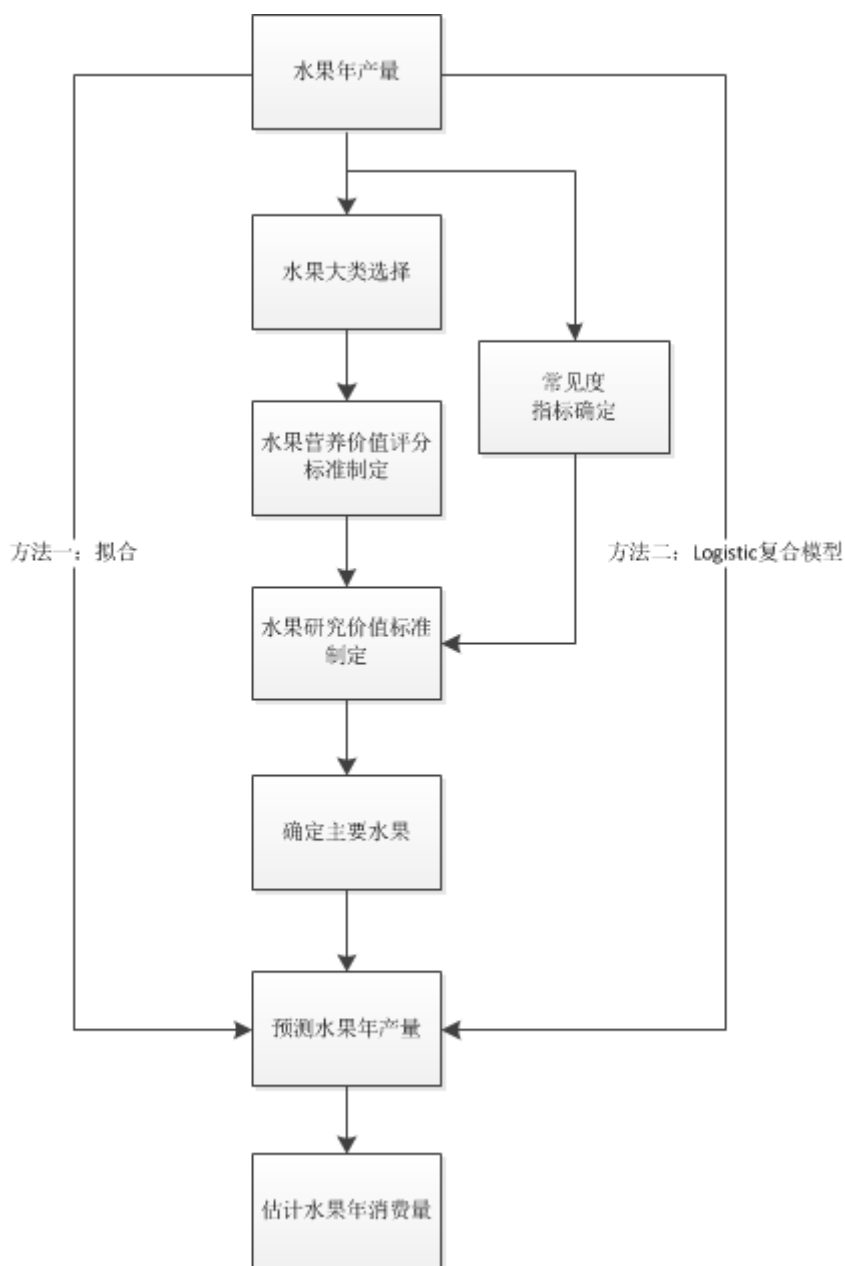


图 1 水果消费量建模技术路线图

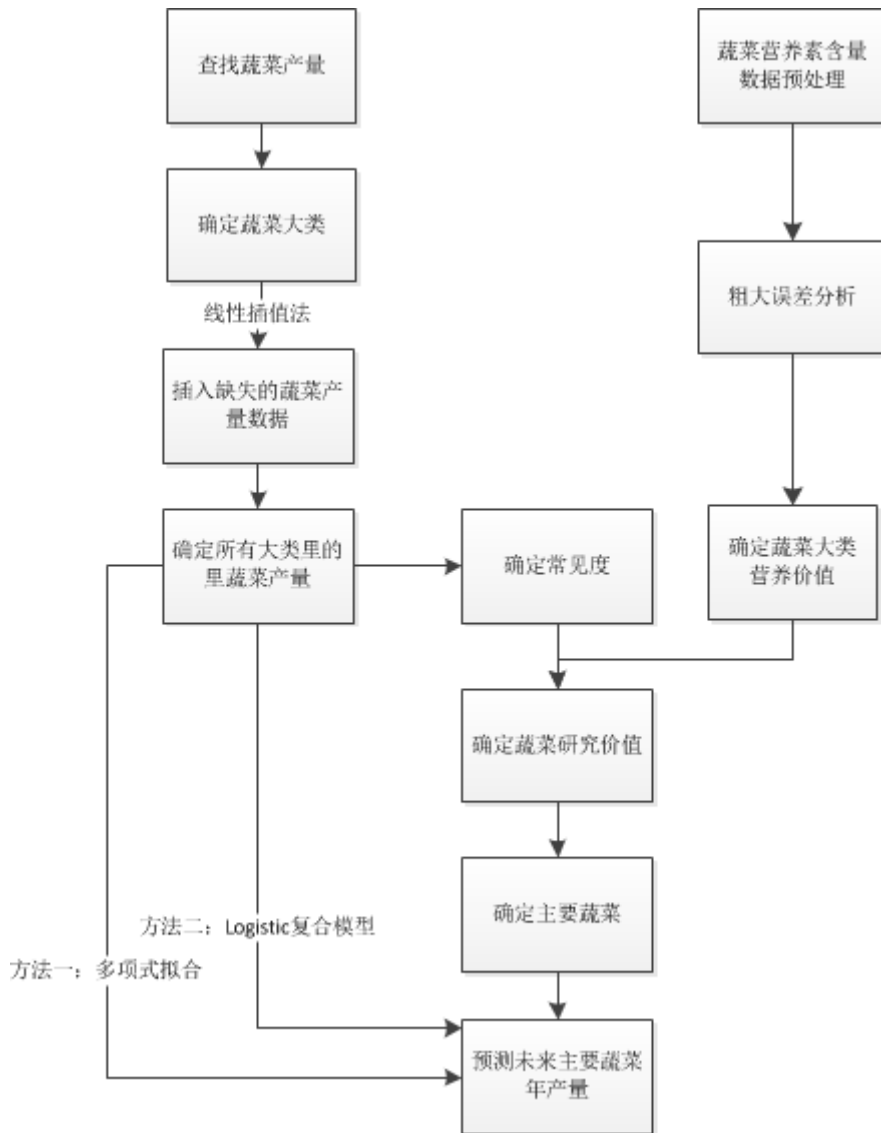


图2 蔬菜消费量建模技术路线图

4.1 确定主要水果

4.1.1 确定研究的营养素范围

水果和蔬菜都是富含营养素的最主要的农产品，主要为人体提供矿物质、维生素、膳食纤维。根据查阅资料可得人体必需从外界摄取的营养素包括维生素 A、A、B1、B2、C、E，矿物质包括钠、钙、铁、锌、硒，并结合题中附件给予的数据，最后我们确定的营养素的范围为：维生素 A，B1，B2，C，E；膳食纤维；矿物质钠，钙，铁，锌，硒。

4.1.2 数据预处理

(1) 原始数据提取

根据题目建议以及文献查阅可知，对主要水果选择既需要考虑水果的营养

素含量又需要考虑到水果的常见度情况。为了对这些情况进行考察，首先从题目附件和国家中华人民共和国农业部种植业管理司查到水果的产量数据，由于某些年份数据不完整，按照数据的完整性初步筛选出我国主要水果品种在2002-2010年的年产量作为研究对象，如表2所示。水果品种为12种，在2002-2010年，这12种水果的总计产量占当年的水果总量的比例约90%，满足研究需求。

表2 常见水果年产量表（单位：万吨）

水果	2002	2003	2004	2005	2006	2007	2008	2009	2010
苹果	1924	2110	2368	2401	2606	2786	2985	3168	3326
橘子	1199	1345	1496	1592	1790	2058	2331	2521	2645
梨	931	980	1064	1132	1199	1290	1354	1426	1506
香蕉	556	590	606	652	690	780	783	883	956
桃子	523	615	701	762	821	905	953	1004	1046
葡萄	448	518	568	579	627	670	715	794	855
大枣	157	172	201	249	305	303	363	425	447
柿子	174	180	200	219	232	257	271	283	
荔枝	152	112	156	144	151	171	151	170	177
龙眼	94	91	102	109	111	117	127	126	131
菠萝	83	82	81	85	89	91	93	104	108
猕猴桃	28	33	41	46	52	58	67	88	107
总年产占比	90%	90%	90%	90%	90%	90%	90%	90%	88%

从题目给出的附件1的常见水果营养成分表中筛选出12种主要水果的相关营养素含量数据，如表3所示。

表3 筛选后水果营养素含量（单位：微克/100克）

水果	A	B1	B2	C	E	钙	铁	钠	锌	硒
苹果	100	0.01	0.03	8	1.46	11	0.1	0.9	0.01	1
橘子	277	0.05	0.04	33	0.45	35	0.2	1.3	1	0.45
梨	100	0.03	0.03	4	1.46	3	0.7	0.7	0.1	0.98
香蕉	56	0.02	0.04	3	0.5	32	0.4	0.4	0.17	0.87
桃子	5	0.01	0.03	9	0.7	12	0.5	1	0.15	0.1
葡萄	5	0.05	0.03	4	0.34	11	0.2	0.5	0.02	0.5
大枣	2	0.06	0.05	297	0.1	16	0.7	7	1.82	1.02
柿子	20	0.02	0.02	30	1.12	9	0.2	0.8	0.08	0.24
荔枝	2	0.02	0.06	36	0.1	6	0.5	1.7	0.17	0.14
龙眼	106	0.01	0.14	43	0	6	0.2	3.9	0.4	0.83
菠萝	33	0.08	0.02	24	0	18	0.5	0.8	0.14	0.24
猕猴桃	66	0.01	0.02	652	1.3	32	0.3	3.3	0.57	0.28

4.1.3 水果营养价值评分标准制定

考虑水果品种对人体的营养价值对12种水果进一步筛选。由于每种水果所含的各营养素并不完全相同，为了综合考虑每种水果各种营养素含量，现采用分段评分的方法，针对各种营养素的含量对每种水果进行打分，然后通过每种水

果的总得分对水果品种营养价值进行评判。

表 4 对水果打分标准

分类	评判标准	得分
不含	$C=0$	0
微含	$0 < C < C_{mid}$	1
富含	$C > C_{mid}$	2

其中，

C_{min} : 某种营养素在各种水果中含量的最小值

C_{mid} : 某种营养素在各种水果中含量的中位数

C_{max} : 某种营养素在各种水果中含量的最大值

根据表 3，可以很容易得出各种营养素的含量最小值、中位数和最大值，见表 5。

表 5 各营养素含量的最小值、中位数、最大值

	A	B1	B2	C	E	钙	铁	钠	锌	硒
C_{min}	2	0.01	0.02	3	0	3	0.1	0.4	0.01	0.1
C_{mid}	44.5	0.02	0.03	27	0.475	11.5	0.35	0.95	0.16	0.475
C_{max}	277	0.08	0.14	652	1.46	35	0.7	7	1.82	1.02

根据表 4 提出的打分标准，代入数据可以得到每种水果对应的营养素的打分，详情见表 6。

表 6 每种水果对应的营养素打分情况

水果	A	B1	B2	C	E	钙	铁	钠	锌	硒	总分
苹果	2	1	2	1	2	1	1	1	1	2	14
橘子	2	2	2	2	1	2	1	2	2	1	17
梨	2	2	2	1	2	1	2	1	1	2	16
香蕉	2	2	2	1	2	2	2	1	2	2	18
桃子	1	1	2	1	2	2	2	2	1	1	15
葡萄	1	2	2	1	1	1	1	1	1	2	13
大枣	1	2	2	2	1	2	2	2	2	2	18
柿子	1	2	1	2	2	1	1	1	1	1	13
荔枝	1	2	2	2	1	1	2	2	2	1	16
龙眼	2	1	2	2	0	1	1	2	2	2	15
菠萝	1	2	1	1	0	2	2	1	1	1	12
猕猴桃	2	1	1	2	2	2	1	2	2	1	16

4.1.4 水果研究价值标准制定

仅仅考虑水果品种的营养价值进行筛选是不合理的。因为如果某水果营养价值高但其产量非常小，那么对于人体来说该种水果就不能为人体提供较多的营养量，因此，考虑水果的产量是非常必要的。所以需要引入另一个指标对水果品种的常见度进行评判，该指标应和水果的产量有一定的关系。基于此，我们定义一个新的指标，常见度 Com 。

$$Com(i, j) = \frac{Pro(i, j)}{Pro(j)} \quad (1)$$

其中 i 表示具体水果种类, j 表示年份, $Pro(i, j)$ 表示第 i 种水果在 j 年的产量, $Pro(j)$ 表示 j 年水果总产量。该指标采用水果的产量数据为依据, 符合客观情况。计算得出 12 种水果的每一年的 $Com(i, j)$ 系数, 并按照每种水果对 2002-2009 年的 $Com(i, j)$ 系统求平均值, 结果如表所示。

表 7 每种水果的常见度

NO.	水果品种	2002	2003	2004	2005	2006	2007	2008	2009	平均值
1	苹果	0.2768	0.2794	0.2820	0.2718	0.2715	0.2648	0.2632	0.2587	0.2710
2	橘子	0.1725	0.1782	0.1782	0.1802	0.1865	0.1956	0.2056	0.2059	0.1878
3	梨	0.1339	0.1298	0.1268	0.1282	0.1249	0.1226	0.1194	0.1165	0.1252
4	香蕉	0.0799	0.0782	0.0721	0.0738	0.0719	0.0741	0.0691	0.0721	0.0739
5	桃子	0.0752	0.0814	0.0835	0.0863	0.0856	0.0860	0.0841	0.0820	0.0830
6	葡萄	0.0644	0.0685	0.0676	0.0656	0.0653	0.0637	0.0631	0.0648	0.0654
7	大枣	0.0226	0.0228	0.0240	0.0282	0.0318	0.0288	0.0320	0.0347	0.0281
8	柿子	0.0250	0.0238	0.0238	0.0247	0.0242	0.0245	0.0239	0.0231	0.0241
9	荔枝	0.0219	0.0149	0.0185	0.0163	0.0157	0.0162	0.0133	0.0138	0.0163
10	龙眼	0.0136	0.0121	0.0121	0.0124	0.0115	0.0111	0.0112	0.0103	0.0118
11	菠萝	0.0119	0.0109	0.0096	0.0096	0.0093	0.0086	0.0082	0.0085	0.0096
12	猕猴桃	0.0041	0.0044	0.0049	0.0052	0.0055	0.0055	0.0059	0.0071	0.0053

4.1.5 确定主要水果

综合考虑水果品种的营养价值和常见度, 定义水果品种研究价值,

$$RVal(i) = Com(i) \cdot NVal(i), \quad (2)$$

$RVal(i)$ 表示第 i 种水果的研究价值, $Com(i)$ 表示第 i 种水果的常见度, $NVal(i)$ 表示第 i 种水果营养价值。根据定义, 计算得出每种水果的研究价值, 如表 8 所示。

表 8 每种水果的研究价值

No.	水果品种	研究价值
1	苹果	3.79
2	橘子	3.19
3	梨	2.00
4	香蕉	1.33
5	桃子	1.25
6	葡萄	0.85
7	大枣	0.51
8	柿子	0.31

9	荔枝	0.26
10	龙眼	0.18
11	菠萝	0.11
12	猕猴桃	0.09

按照每种水果的研究价值从大到小排序，挑选前 8 位的水果品种，见表 9。

表 9 主要水果和其研究价值

苹果	橘子	梨	香蕉	桃子	葡萄	大枣	柿子
3.79	3.19	2.00	1.33	1.25	0.85	0.51	0.31

4.2 水果消费量建模

4.2.1 基于相关分析的变量分析

相关分析是研究变量之间密切程度的一种统计方法。本节通过相关分析可以了解水果产量与年份间关系的密切程度，为预测水果每年产量做出准备。

任意两个变量间的皮尔森相关系数 ρ 可由公式 (3) 计算得到。

$$\rho = \frac{\sum_{k=1}^n (X_{ik} - \bar{X}_i)(X_{jk} - \bar{X}_j)}{\sqrt{\sum_{k=1}^n (X_{ik} - \bar{X}_i)^2 \sum_{k=1}^n (X_{jk} - \bar{X}_j)^2}} \quad (3)$$

其中， \bar{X} 代表单个数据的均值； X_{ik} 代表第 k 个变量第 i 个数据的数值； X_{jk} 代表第 k 个变量第 j 个数据的数值； ρ 为相关系数。

利用上述筛选的水果（苹果、梨、橘子、香蕉、桃子、葡萄、大枣、柿子）和查找得到的 2002-2009 年年产量数据，在 SPSS 软件中进行相关分析，可得到水果年产量和年份的相关系数，如表 10 所示。

表 10 水果年产量和年份相关系数

水果	苹果	梨	橘子	香蕉	桃子	葡萄	大枣	柿子
相关系数	0.9954	0.9989	0.9988	0.9787	0.9956	0.9896	0.9851	0.9951

一般认为当皮尔森相关系数的绝对值大于 0.8 时，两个变量之间具有很强相关性，并且越接近 1 相关性越强。根据这一准则，可以看出以上水果中每年的年产量和年份的相关系数都约为 1，表明所选择的每种水果的年产量和年份之间都有强烈的相关关系。

4.2.2 水果年消费量估计及发展趋势

(1) 线性回归法

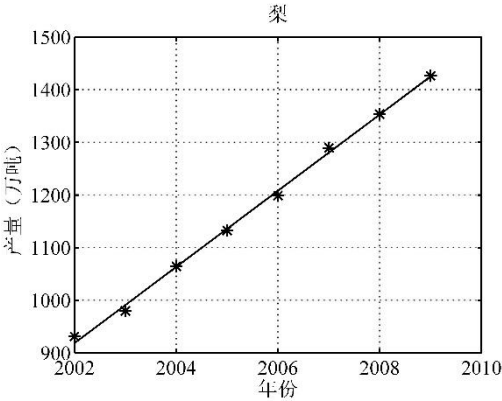
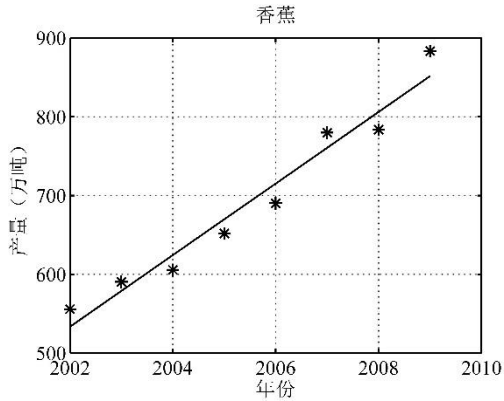
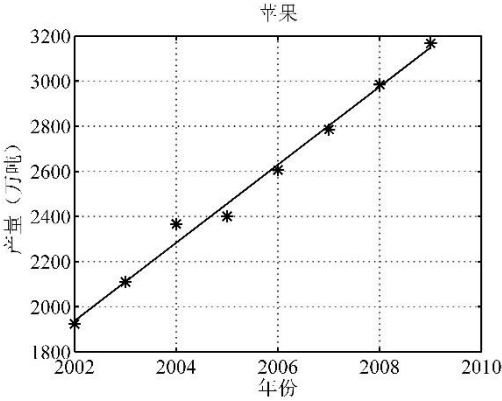
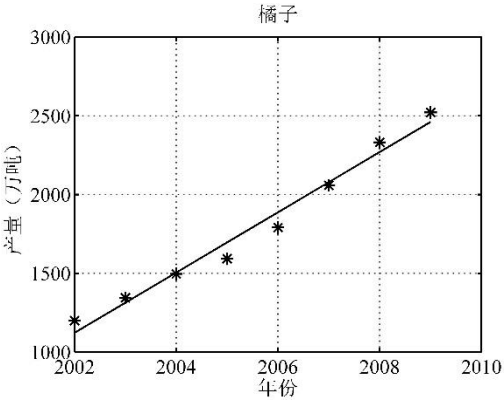
线性回归是利用数理统计中的回归分析，来确定两种或两种以上变量间相互依赖的定量关系的一种统计分析方法。当自变量和因变量之间存在较强线性关系时，便可以线性回归分析进行预测分析。由上一小节可以得到主要水果的每年年产量和年份存在强烈的线性相关关系，因此在本节中首先利用

2002-2009 年的年产量数据得出每种水果年产量和年份的函数，进而可以得到水果 2010 年的预测年产量，用此预测的产量与 2010 年的真是产量对比，来分析精准度。

利用 SPSS 软件对主要水果进行回归，回归方程系数见表 11，各水果的回归图见图 3。

表 11 回归方程系数

水果	a1	a0
苹果	173	-344610
橘子	191	-381864
梨	72	-143974
香蕉	45	-90505
桃子	68	-136048
葡萄	45	-89267
大枣	38	-75932
柿子	17	-33431



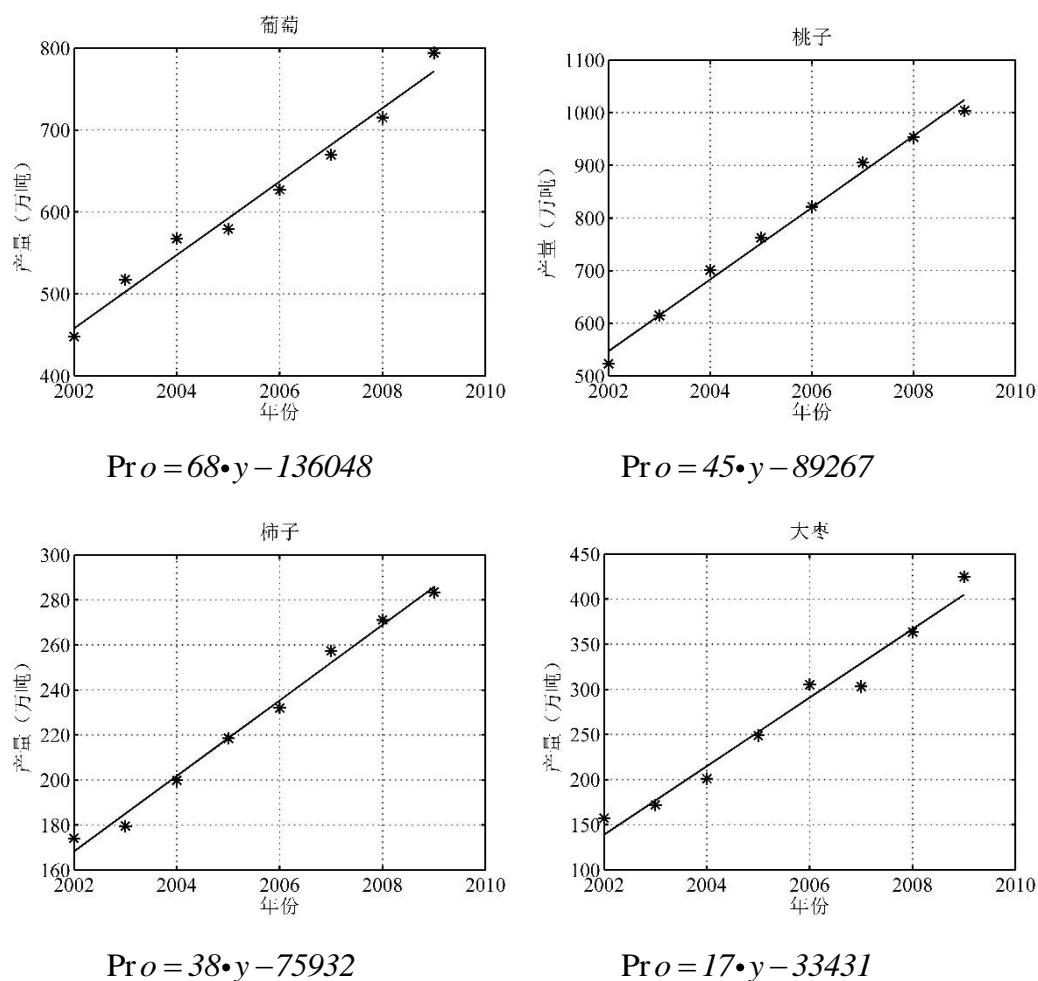


图3 各种水果的线性回归图

通过线性回归的结果分别计算 2010 年苹果、橘子、梨、香蕉、桃子、葡萄和大枣 7 种水果的产量，并与所查到的实际数据对比，结果如表 12 所示。

表 12 预测值与实际值对比

水果	2010 年原值	2010 年预测值	绝对误差	相对误差
苹果	3326	3322	4	0.12%
橘子	2645	2652	7	0.27%
梨	1506	1498	8	0.54%
香蕉	956	897	59	6.16%
桃子	1046	1093	47	4.51%
葡萄	855	816	38	4.49%
大枣	447	443	4	0.87%

误差最大为香蕉的产量，相对误差为 6.16%，其余的水果相对误差非常小，基本可认为预测非常精准。

由上节进行的预测回归，可以得出各种水果的每年产量模型，见表 13。

表 13 主要水果的年产量函数

水果	年产量函数
苹果	$Pro = 173 \cdot y - 344610$

橘子	$Pro = 191 \cdot y - 381864$
梨	$Pro = 72 \cdot y - 143974$
香蕉	$Pro = 45 \cdot y - 90505$
桃子	$Pro = 68 \cdot y - 136048$
葡萄	$Pro = 45 \cdot y - 89267$
大枣	$Pro = 38 \cdot y - 75932$
柿子	$Pro = 17 \cdot y - 33431$

根据题中附件 4，可认为损耗率是浪费的比例，设定田间地头到大市场损耗率 $w1$ ，大市场到零售市场损耗率 $w2$ ，零售市场到餐桌的损耗率 $w3$ ，损耗率见表 14。

表 14 损耗率

水果	$w1$	$w2$	$w3$
苹果	0.03	0.01	0.01
橘子	0.10	0.04	0.06
梨	0.08	0.03	0.03
香蕉	0.21	0.09	0.12
桃子	0.08	0.03	0.03
葡萄	0.2	0.08	0.15
大枣	0.05	0.01	0.03
柿子	0.10	0.04	0.10

因此给出消费量 Con 的估计模型为： $Con = Pro(1 - w1)(1 - w2)(1 - w3)$ 。(4)

根据前文预测的水果产量函数，分别代入之后可以得出每种水果的消费量估计函数，见表 15。

表 15 主要水果的消费量估计

水果种类	消费量函数
苹果	$Con = (173 \cdot y - 344610)(1 - w1)(1 - w2)(1 - w3)$
梨	$Con = (72 \cdot y - 143974)(1 - w1)(1 - w2)(1 - w3)$
桃子	$Con = (68 \cdot y - 136048)(1 - w1)(1 - w2)(1 - w3)$
葡萄	$Con = (45 \cdot y - 89267)(1 - w1)(1 - w2)(1 - w3)$
香蕉	$Con = (45 \cdot y - 90505)(1 - w1)(1 - w2)(1 - w3)$

橘子	$Con = (191 \cdot y - 381864)(1 - w1)(1 - w2)(1 - w3)$
柿子	$Con = (17 \cdot y - 33431)(1 - w1)(1 - w2)(1 - w3)$
大枣	$Con = (38 \cdot y - 75932)(1 - w1)(1 - w2)(1 - w3)$

由上表可知，每种水果的年消费量呈现线性增长形势，其中苹果橘子的增长速率较大，其余增长相对平缓。按照所有水果的消费量估计模型，为了更客观地描述其发展趋势，首先我们估计出从 2014-2025 年水果的消费量，再画出相应的折线图来描述。见表 16 和图 4。

表 16 2014-2025 年水果消费量

水果	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
苹果	3624	3789	3953	4117	4282	4446	4611	4775	4940	5104	5269	5433
梨	895	957	1020	1082	1144	1207	1269	1331	1394	1456	1518	1581
桃子	783	841	900	959	1018	1077	1136	1195	1253	1312	1371	1430
葡萄	853	881	909	937	965	993	1022	1050	1078	1106	1134	1162
香蕉	79	108	136	164	193	221	250	278	307	335	364	392
橘子	2282	2437	2592	2748	2903	3058	3213	3368	3523	3678	3833	3989
柿子	628	641	654	667	680	694	707	720	733	746	760	773
大枣	547	582	617	651	686	721	755	790	825	859	894	929

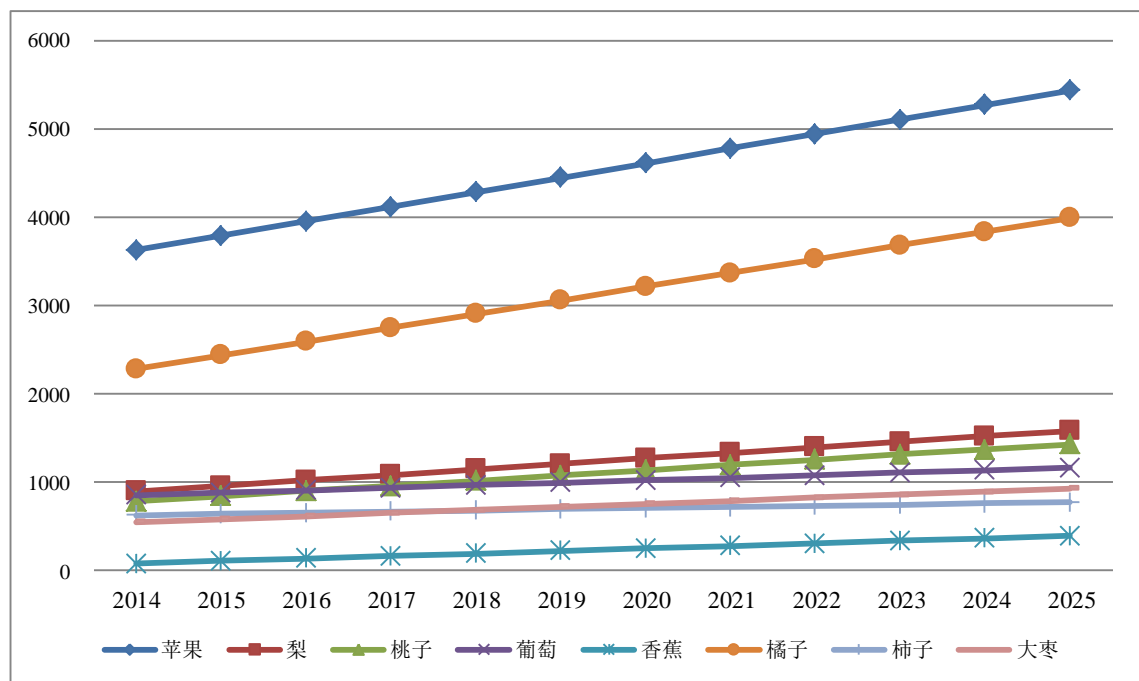


图 4 2014-2025 年水果消费量预测

(2) Logistic 模型法

1837 年荷兰生物数学家 Verhulst 提出生物群体增长的逻辑律，可表示为方程：

$$\frac{dN(t)}{dt} = \alpha N(t) - \beta N(t)^2 \quad (5)$$

其中 $\alpha > 0$, $\beta > 0$, 均为常数, 且 α 和 β 相比是个比较微小的数。后来, 人们又称上述非线性方程为生物种群增长的 Logistic 模型, 或者阻滞增长模型, 这一模型是在下列假设的基础上建立的:

(a) 假设有一个环境条件允许的种群数量的最大值, 这个数值称为环境容纳量或者负载量, 通常用 M 表示。当种群数量达到 M 时, 种群将不再增长, 即

$$\frac{dN(t)}{dt} = 0。$$

(b) 假设环境条件对种群的阻滞作用随着种群的密度的增加而按比例增加。根据阻滞增长模型, 果蔬产量预测可表示为微分方程初始问题:

$$\begin{cases} \frac{dN(t)}{dt} = [\alpha - \beta N(t)]N(t) \\ N(t_0) = N_0 \end{cases} \quad (6)$$

其中 $P(t) = \alpha - \beta N(t)$ 是产量增长率; α 为产量的自然增长率; β 为产量系数; $N(t)$ 为 t 时刻的总产量; $N(t_0)$ 为初始时刻 t_0 的总产量。

求解此微分方程的初值问题得:

$$N(t) = N_0 \cdot \exp\left[-\left(bt_0 + \frac{a}{2}t_0^2\right)\right] \cdot \exp\left(bt + \frac{a}{2}t^2\right) \quad (7)$$

由于模型 (7) 是由两个模型 (5) 和 (6) 复合而成的, 所以称它为复合模型, 即 Logistic 产量预测复合模型。

Logistic 模型考虑到了自然资源、环境条件等因素对产量的增长起着阻滞作用, 并且随着人们对食物选择种类的不断增加, 阻滞作用将越来越大。Logistic 模型是对马尔萨斯指数增长模型的基本假设进行修改后得到的。

研究 Logistic 后, 我们在 Logistic 模型的基础上建立一种预测水果产量的复合模型; 然后, 利用此模型来预测未来水果产量。由于 Logistic 模型自身的优越性, 我们可以预测较长时间范围的水果产量。

以苹果消费量数据为例, 通过 Logistic 模型对苹果未来消费量计算进行预测, 其他品种水果可以以此类推。

表 17 2002-2012 年苹果产量

序号	年份	产量	增长量	年增长率
1	2002	1924	-	-
2	2003	2110	186	0.0967
3	2004	2368	257	0.1220
4	2005	2401	34	0.0142
5	2006	2606	205	0.0853
6	2007	2786	180	0.0691
7	2008	2985	199	0.0713
8	2009	3168	183	0.0615
9	2010	3326	158	0.0500

假设产量增长率与年份之间呈线性关系，即

$$r(y) = ay + b, \quad a, b \text{ 是参数}, \quad ab < 0$$

选取 2003-2010 年增长率与年份的数据点，构造一元线性回归，拟合结果如图 5 所示。

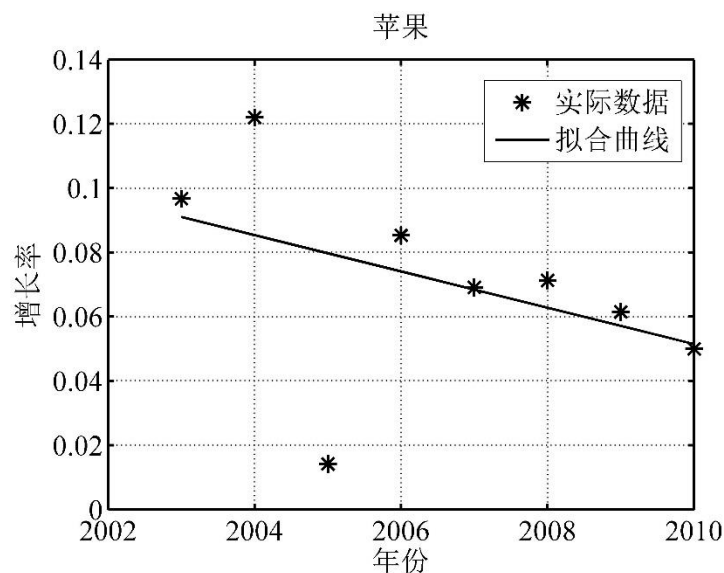


图 5 苹果的年增长率与年份拟合

忽略拟合误差，于是得到产量增长率模型：

$$r = -0.00565y + 11.40978$$

代入 Logistic 复合模型，并取初值 $Pro(2010) = 3326$ ，可以预测出苹果的产量在 2011-2030 年走势图，如图 6 所示。对比 2002-2010 年的实际数据，可以发现拟合出的曲线与实际数据的误差很小，拟合效果非常好，说明模型适合 2002-2010 年苹果的产量状况。

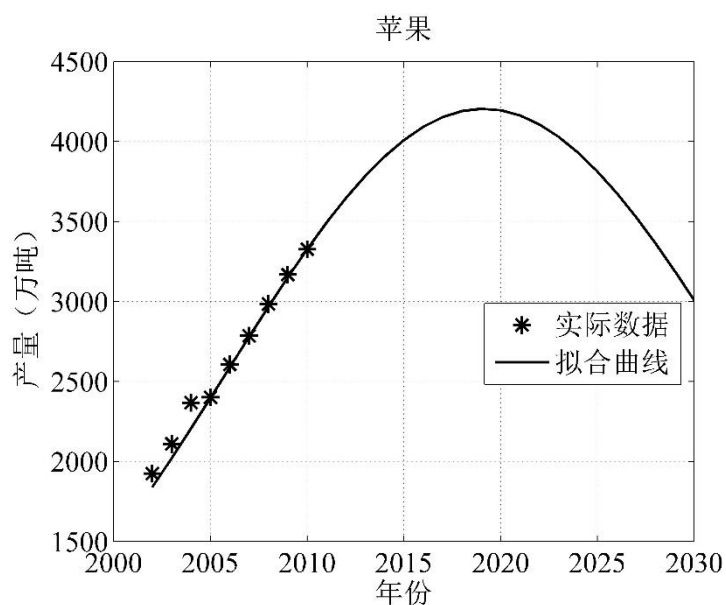


图 6 苹果的年产量走势图

通过拟合曲线可以预测出 2011-2030 年的苹果产量，见表 18。产量在 2019 年达到最大值，约为 4205 万吨，在 2019 年之前，产量持续增长，2019 年之后产量逐渐下降。其他几种主要水果产量数据采用同样的方法建模，得到图形如图 7 所示。在这些基础上，再预测 2015-2030 年水果产量，见表 19。

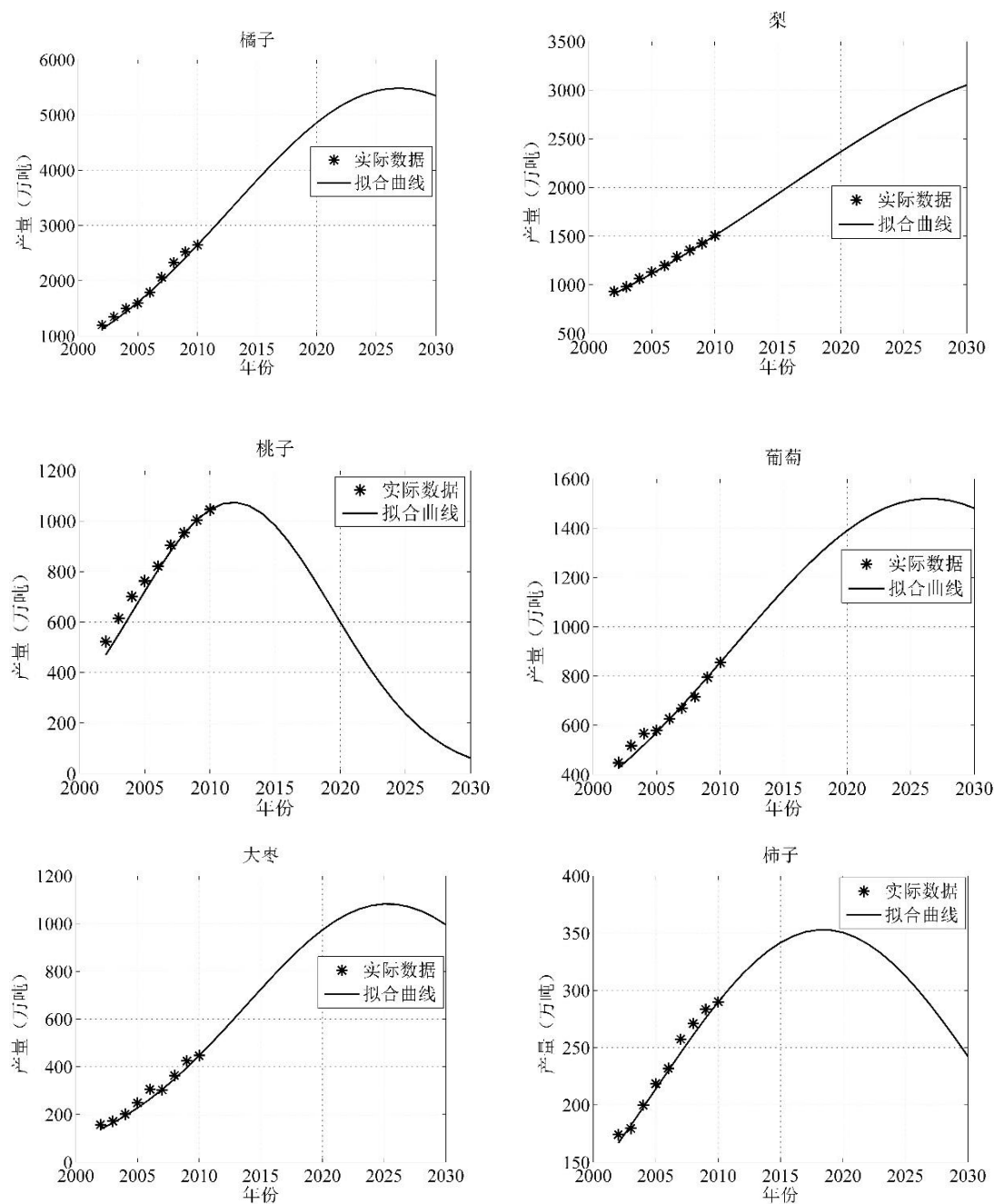


图 7 各种水果年产量走势图

表 18 水果产量最大值及达到最大值的年份

	苹果	橘子	梨	香蕉	桃子	葡萄	大枣	柿子
最大值 (万吨)	4205	5485	2030	2030	982	1520	1083	353
达到最大值的时间	2019	2027	2643	12671	2015	2026	2025	2018

表 19 各种水果在 2015-2030 年的产量 (单位: 万吨)

年份	苹果	橘子	梨	香蕉	桃子	葡萄	大枣	柿子
2015	4009	3824	1935	1633	982	1150	725	342

2016	4092	4053	2022	1850	921	1205	781	347
2017	4152	4274	2110	2109	849	1257	835	351
2018	4190	4485	2196	2419	769	1306	886	353
2019	4205	4681	2282	2791	685	1351	932	353
2020	4195	4862	2367	3240	599	1391	974	351
2021	4163	5023	2449	3784	515	1427	1010	346
2022	4107	5163	2530	4445	435	1457	1040	340
2023	4029	5280	2608	5254	362	1482	1062	333
2024	3930	5372	2683	6247	295	1501	1076	323
2025	3812	5437	2755	7472	237	1514	1083	313
2026	3677	5475	2823	8992	187	1520	1081	300
2027	3526	5485	2888	10887	145	1520	1071	287
2028	3363	5467	2947	13260	110	1513	1053	273
2029	3189	5422	3003	16248	83	1500	1028	258
2030	3007	5349	3053	20028	61	1481	995	242

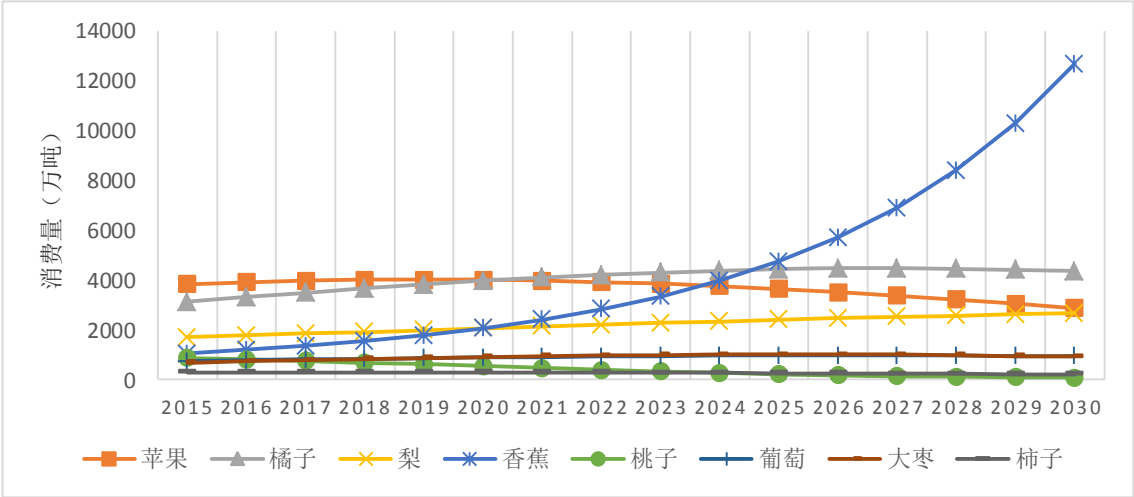


图7 水果消费量图

通过损耗率公式（4）转化成消费量量如表 20 所示。（灰色框表示在该年达到最大值）

表 20 水果 2015-2030 消费量预测值

年份	苹果	橘子	梨	香蕉	桃子	葡萄	大枣	柿子
2015	3811	3105	1675	1033	850	719	662	266
2016	3890	3292	1751	1170	797	754	713	270
2017	3948	3472	1826	1334	735	786	762	273
2018	3984	3642	1901	1530	666	817	808	274
2019	3997	3802	1976	1766	593	845	851	274
2020	3989	3948	2049	2050	519	870	889	273
2021	3957	4079	2120	2394	446	893	922	269
2022	3904	4193	2190	2812	377	912	948	265
2023	3830	4288	2258	3324	313	927	969	259
2024	3736	4363	2323	3952	256	939	982	252

2025	3624	4416	2385	4727	205	947	988	243
2026	3495	4447	2444	5689	162	951	986	234
2027	3352	4455	2500	6887	125	951	977	223
2028	3197	4440	2551	8389	96	947	961	212
2029	3032	4403	2599	10279	72	938	938	201
2030	2859	4344	2643	12671	53	927	908	189

4.3 确定主要蔬菜

4.3.1 数据预处理

(1) 产量数据提取及确定常见度取值

从中华人民共和国农业部种植业管理司查找蔬菜产量数据，由于某些年份数据不完整，按照数据的完整性初步筛选出我国主要蔬菜品种在 1990-2009 年的年产量作为研究对象，如表 21 所示。蔬菜品种为 9 种常见蔬菜，表格空白处缺失相应的数据，为了之后的研究，现采用线性插值的方法补全所缺数据。

表 21 常见蔬菜产量

年份	萝卜	胡萝卜	蒜苗	茄子	黄瓜	大白菜	芹菜	菠菜
1990	1715	296	439	511	720	3555	511	281
1991	1287	303	497	579	758	3578	543	321
1992	1049	224	437	558	975	2673	507	293
1993	1015	348	539	653	1075	2916	610	330
1994	1216	305	494	759	1256	3034	617	324
1995								
1996								
1997								
1998								
1999								
2000								
2001								
2002								
2003	3881	1312	1556	2119	3551	10197	1796	1574
2004	3833	1330	1567		3655	10346	1919	1567
2005	3935	1331	1654	2263	3817	10308	1951	1618
2006	4003	1443	1832	2247	4040	10506	2068	1627
2007								
2008	3955	1477	1828	2372	4219	10546	1976	1663
2009	4080	1506	1790	2589	4420	10587	2081	1754

在 Matlab 2010 中调用线性插值数，根据上表中已有的产量数据，对 9 种蔬菜的缺失产量数据的年份进行线性插值计算，得到如图 8 所示的结果。图中，黑色点表示原有数据，绿色点表示插值数据。

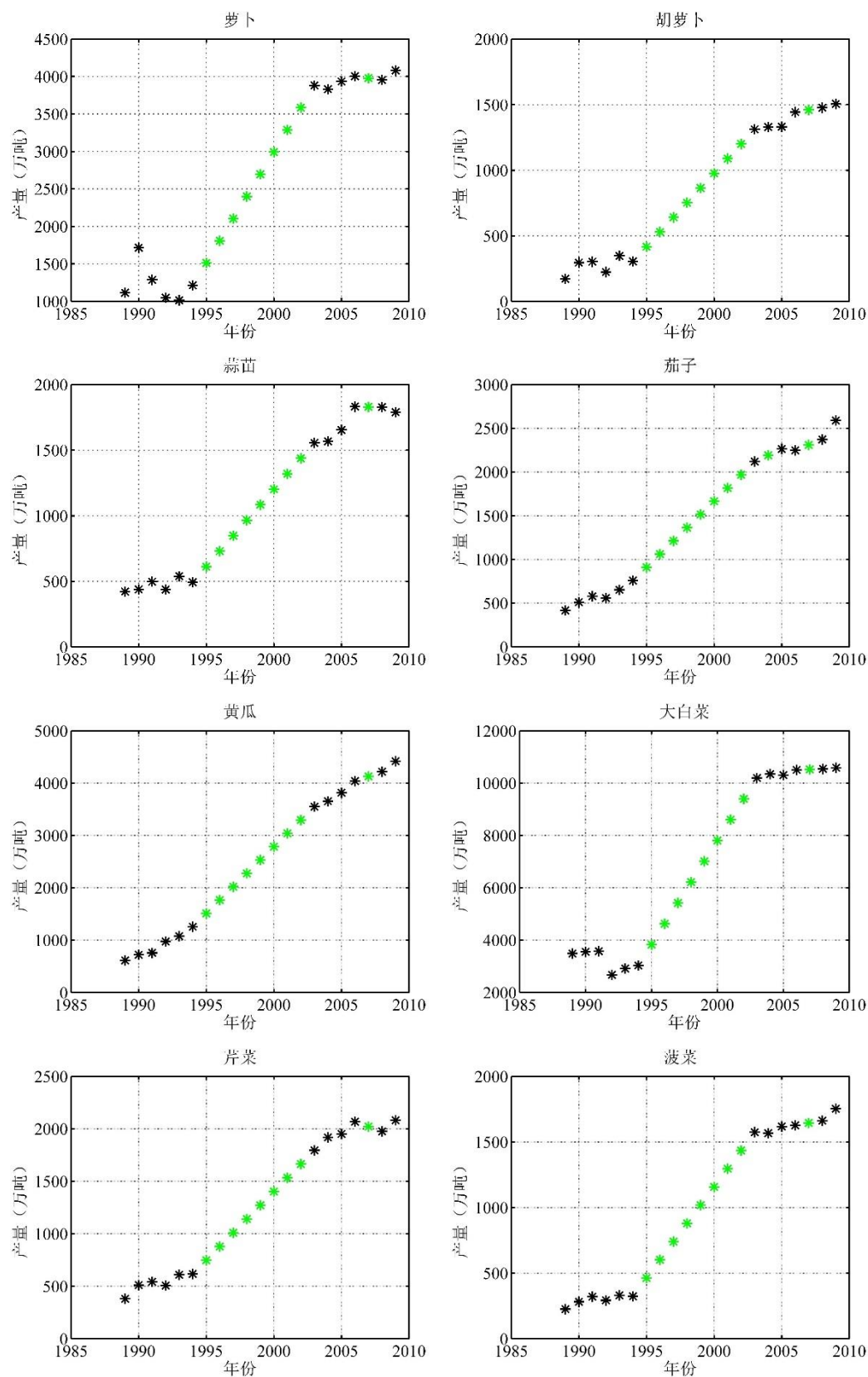


图8 蔬菜插值图

插值得到的 1990-2009 年完整的蔬菜产量数据如表 22 所示。

表 22 插值后的蔬菜产量

年份	萝卜	胡萝卜	蒜苗	茄子	黄瓜	大白菜	芹菜	菠菜
----	----	-----	----	----	----	-----	----	----

1990	1715	296	439	511	720	3555	511	281
1991	1287	303	497	579	758	3578	543	321
1992	1049	224	437	558	975	2673	507	293
1993	1015	348	539	653	1075	2916	610	330
1994	1216	305	494	759	1256	3034	617	324
1995	1512	417	612	910	1511	3830	748	463
1996	1809	529	730	1061	1766	4626	879	602
1997	2105	641	848	1212	2021	5422	1010	741
1998	2401	753	966	1363	2276	6218	1141	880
1999	2697	865	1084	1515	2531	7014	1272	1018
2000	2993	977	1202	1666	2786	7810	1403	1157
2001	3289	1089	1320	1817	3041	8606	1534	1296
2002	3585	1201	1438	1968	3296	9401	1665	1435
2003	3881	1312	1556	2119	3551	10197	1796	1574
2004	3833	1330	1567	2191	3655	10346	1919	1567
2005	3935	1331	1654	2263	3817	10308	1951	1618
2006	4003	1443	1832	2247	4040	10506	2068	1627
2007	3979	1460	1830	2310	4130	10526	2022	1645
2008	3955	1477	1828	2372	4219	10546	1976	1663
2009	4080	1506	1790	2589	4420	10587	2081	1754

为了从以上 9 中蔬菜中筛选出具有研究价值的蔬菜品种，综合考虑蔬菜品种的营养价值 $NVal$ 和常见度系数 Com 。由于缺乏相关数据，忽略以上 9 种蔬菜以外的其他蔬菜品种。和水果处理类似，由于表格中很多数据均为插值所得，并不完全准确，因此挑选 2009 年完整的数据计算某种蔬菜品种的常见度系数。

$$Com(i, j) = \frac{Pro(i, j)}{Pro(j)}$$

经计算可得蔬菜的常见度系数，见表 23

表 23 九种蔬菜的常见度

蔬菜	常见度 Com
大白菜	0.116818
萝卜	0.045017
黄瓜	0.048774
菠菜	0.019351
蒜苗	0.01975
茄子	0.028561
芹菜	0.022966
胡萝卜	0.016614

4.3.2 粗大误差分析

(1) 营养素含量数据预处理

为避免数据错误，避免引入误差，首先必须对数据进行预处理。从附件 1

常见蔬菜营养成分表中初步筛选出所有蔬菜各营养素含量，如表 24 所示。

表 24 蔬菜各营养素含量

食物名称	可食部	膳食纤维	维生素 A	维生素 B1	维生素 B2	维生素 C	维生素 E	钠	钙	铁	锌	硒
100g	%	克	微克	微克	毫克	毫克	毫克	毫克	毫克	毫克	毫克	微克
胡萝卜	96	1.1	688	0.04	0.03	13	0.41	71.4	32	1	0.23	0.63
萝卜	94	0.6	3	0.03	0.06	18	1	60	56	0.3	0.13	0.6
竹笋	63	1.8	5	0.08	0.08	5	0.05	0.4	9	0.5	0.33	0.04
大白菜	92	0.6	42	0.06	0.07	47	0.92	89.3	69	0.5	0.21	0.33
菠菜	89	1.7	487	0.04	0.11	32	1.74	85.2	66	2.9	0.85	0.97
菜花	82	1.2	5	0.03	0.08	61	0.43	31.6	23	1.1	0.38	0.73
韭菜	90	1.4	235	0.02	0.09	24	0.96	8.1	42	1.6	0.43	1.38
芹菜	66	1.4	10	0.01	0.08	12	2.21	73.8	48	0.8	0.46	0.5
生菜	94	0.7	298	0.03	0.06	13	1.02	32.8	34	0.9	0.27	1.05
蒜苗	82	1.8	47	0.11	0.08	35	0.81	5.1	29	1.4	0.46	1.24
油菜	87	1.1	103	0.04	0.11	36	0.88	55.8	108	1.2	0.33	0.79
圆白菜	86	1	12	0.03	0.03	40	0.5	27.2	49	0.6	0.25	0.96
冬瓜	80	0.7	13	0.01	0.01	18	0.08	1.8	19	0.2	0.07	0.22
西红柿	97	0.5	92	0.03	0.03	19	0.57	5	10	0.4	0.13	0.15
青椒	84	2.1	57	0.03	0.04	62	0.88	2.2	15	0.7	0.22	0.62
茄子	93	1.3	8	0.02	0.04	5	1013	5.4	24	0.5	0.23	0.48
黄瓜	92	0.5	15	0.02	0.03	9	0.46	4.9	24	0.5	0.18	0.38
苦瓜	81	1.4	17	0.03	0.03	56	0.85	2.5	14	0.7	0.36	0.36
南瓜	85	0.8	148	0.03	0.04	8	0.36	0.8	16	0.4	0.14	0.46
丝瓜	83	0.6	15	0.02	0.04	5	0.22	2.6	14	0.4	0.21	0.86
土豆	94	0.7	5	0.08	0.04	27	0.34	2.7	8	0.8	0.37	0.78
榨菜	100	2.1	83	0.03	0.06	2	0	4253	155	3.9	0.63	1.93
蘑菇	100	21	273	0.1	1.1	5	6.18	23.3	127	10	6.29	39.18
木耳	100	29.9	17	0.17	0.44	0	11.34	48.5	247	97	3.18	3.72
香菇	95	31.6	3	0.19	1.26	5	0.66	11.2	83	11	8.57	6.42

(2) 粗大误差分析

粗大误差或过失误差是样本中的个别值，其数值明显偏离其他样本值，亦称作异常值。在数据处理时，如果对混有离群值的测量数据未予剔除，则会造成分析结果精度偏低；反之如果将正常数据错误地当作离群值予以剔除，就会减少有用的测量信息。因此对测量数据中离群值的合理判别与处理，是获得正确分析结果的前提条件。

本文采用学生化残差对原始数据中的粗大误差进行分析，基本概念如下：

设对某量进行 n 次独立测量，得到的测量列 X 为：

$$X = \{x_1, x_2, \dots, x_i, \dots, x_n\} \quad (i=1, 2, \dots, n) \quad (8)$$

它的算术平均值 \bar{x} 及其残差 v_i 分别为：

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (9)$$

$$v_i = x_i - \bar{x} \quad (10)$$

定义 1: 残差与样本标准差之比为样本学生化残差, 即:

$$y_i = \frac{v_i}{s} \quad (11)$$

式中: s 为样本标准差, 其值为:

$$s = \sqrt{\frac{\sum_{i=1}^n v_i^2}{n}} \quad (12)$$

定义 2: 残差的绝对值与样本标准差之比为样本学生化残差的绝对值, 即:

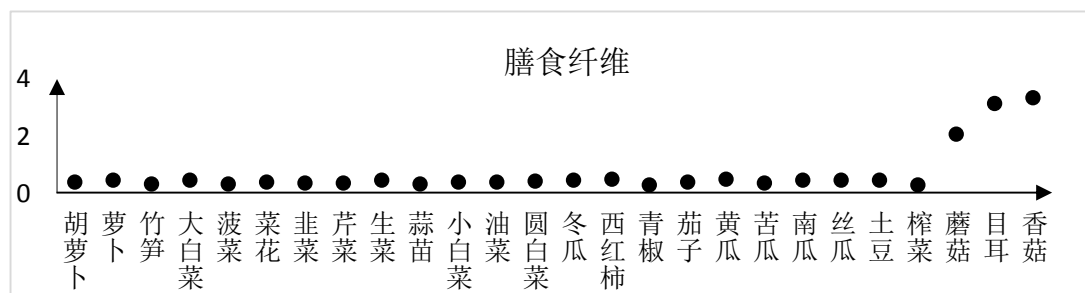
$$|y_i| = \frac{|v_i|}{s} \quad (13)$$

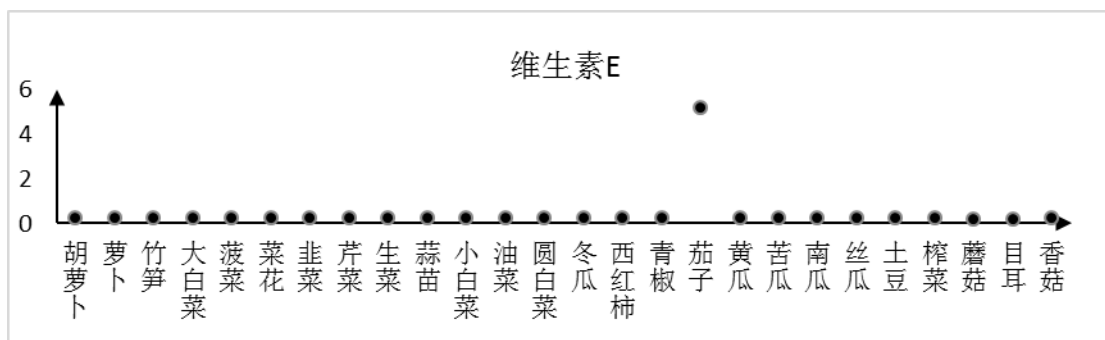
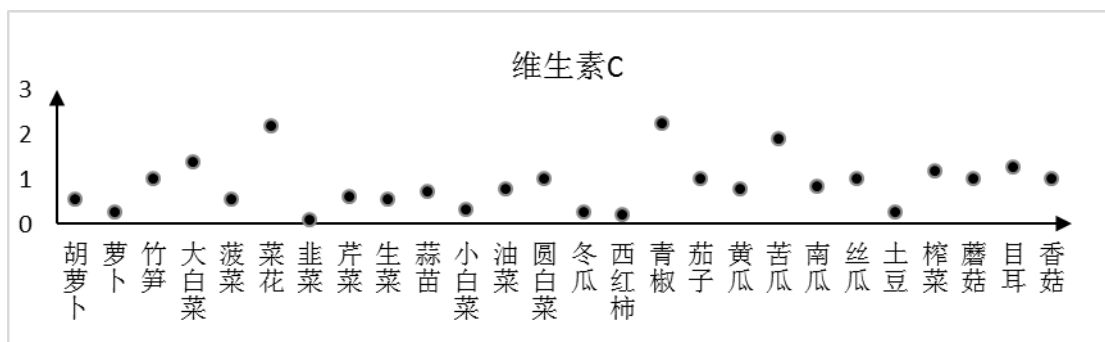
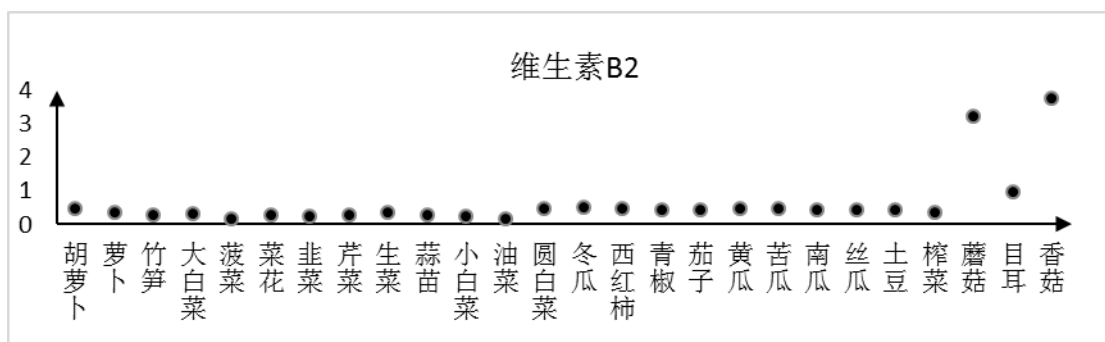
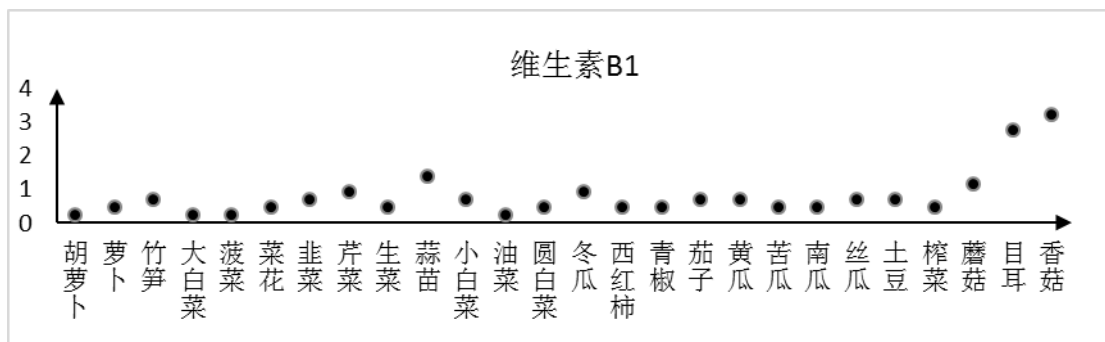
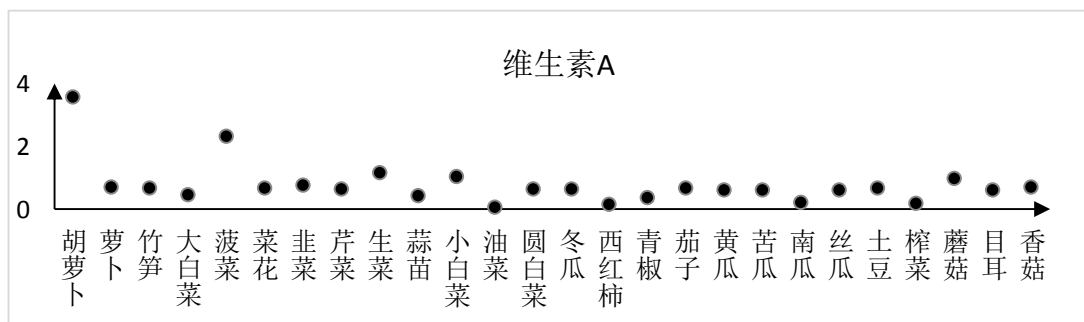
这里所谓的“学生化”主要是针对普通残差不具备方差齐性而进行的一种标准化处理, 使之期望为 0, 方差为 1, 从而便于比较。

根据上述学生化残差定义, 对各蔬菜品种的营养素含量数据进行计算, 结果如表 25 所示。当学生化残差的绝对值大于 5 时认为该数据可能存在错误, 需要校验。从图 9 中可以看出以下数据需要验证, 并参照中国食物成分表 (2002 版) 查找标准值, 可以得出以下结果。

表 25 可能存在错误的数据验证

蔬菜	营养素种类	学生化残差的绝对值	原值	标准值	验证结果
茄子	维生素 E	5.18838	1013	1.13	原值为错误值
榨菜	钠	5.185286148	4253	4253	正确
木耳	铁	5.135058386	97	97	正确
蘑菇	硒	5.111218425	39.18	39.18	正确





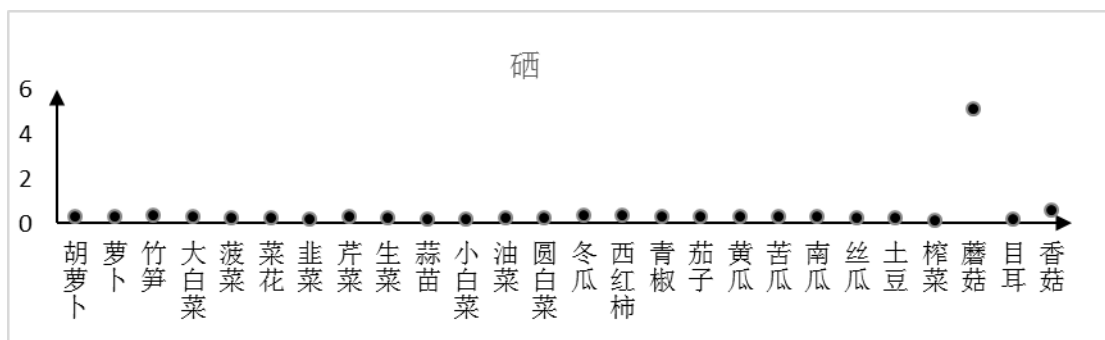
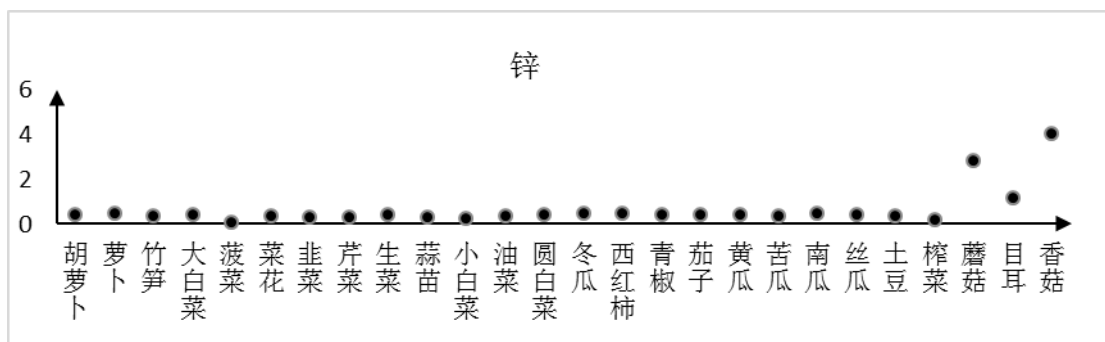
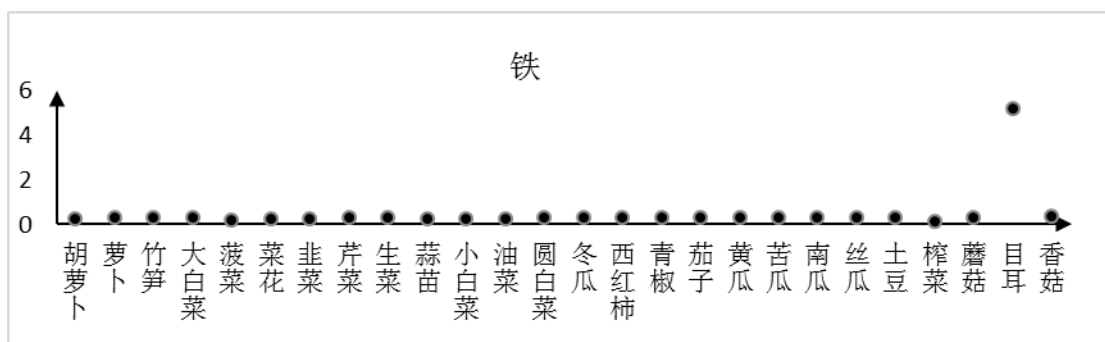
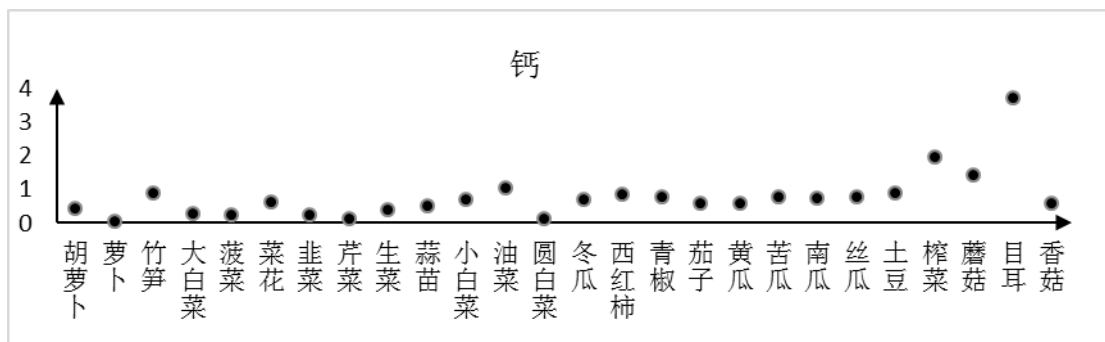
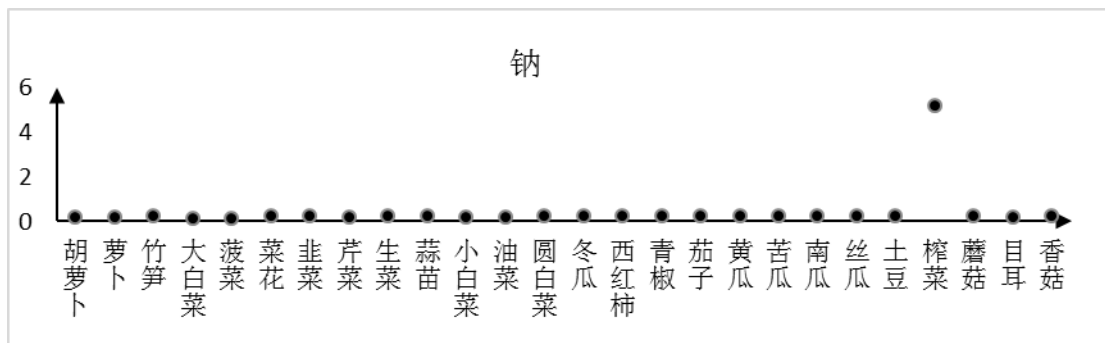


图9 各种营养素粗大误差分析图

经上述表格及图可以得知,可能存在错误的数值是茄子的维生素含量,原表中数据为1013,经查中国食物成分表(2002版)可知数据出现错误,正确数据为1.13,其余数据未查出错误。

表中各种营养物含量均为100g蔬菜中的总量,为了之后研究居民摄入的营养素状况,应该去除蔬菜不可食部分的含量。为了简化计算,假设每种蔬菜可食部分和不可食部分所含营养素含量相同,即将每种营养素的含量乘以相应的蔬菜可食百分比,得每100g蔬菜可食的营养素含量。

提取出所研究的9种蔬菜的相关数据,如表26。

表26 九种蔬菜含量数据

食物名称	膳食纤维	维生素A	维生素B1	维生素B2	维生素C	维生素E	钠	钙	铁	锌	硒
胡萝卜	1.056	660.48	0.0384	0.0288	12.48	0.3936	68.544	30.72	0.96	0.2208	0.6048
萝卜	0.564	2.82	0.0282	0.0564	16.92	0.94	56.4	52.64	0.282	0.1222	0.564
大白菜	0.552	38.64	0.0552	0.0644	43.24	0.8464	82.156	63.48	0.46	0.1932	0.3036
菠菜	1.513	433.43	0.0356	0.0979	28.48	1.5486	75.828	58.74	2.581	0.7565	0.8633
芹菜	0.924	6.6	0.0066	0.0528	7.92	1.4586	48.708	31.68	0.528	0.3036	0.33
蒜苗	1.476	38.54	0.0902	0.0656	28.7	0.6642	4.182	23.78	1.148	0.3772	1.0168
茄子	1.209	7.44	0.0186	0.0372	4.65	1.0509	5.022	22.32	0.465	0.2139	0.4464
黄瓜	0.46	13.8	0.0184	0.0276	8.28	0.4232	4.508	22.08	0.46	0.1656	0.3496

4.3.3 运用分段评分法确定主要蔬菜

选取主要的蔬菜进行研究,需要考虑蔬菜品种对人体的重要性,即研究价值,由于每种蔬菜所含的各营养素并不均衡,为了综合考虑每种蔬菜各个营养素含量,现采用分段评分的方法,针对各个营养素的含量对每种蔬菜进行打分,然后通过每种蔬菜的总得分对蔬菜品种研究价值进行从大到小排序,最后选取总得分较高的总计产量超过蔬菜总产量90%的蔬菜种类。

蔬菜营养评分标准和水果一致,见表27。运用评分标准,可以得出蔬菜的对应评分表,见表5-25。

表27 蔬菜评分表

食物名称	膳食纤维	维生素A	维生素B1	维生素B2	维生素C	维生素E	钠	钙	铁	锌	硒	总分
胡萝卜	2	2	2	1	1	1	2	1	2	2	2	18
萝卜	1	1	2	2	2	2	2	2	1	1	2	18
大白菜	1	2	2	2	2	2	2	2	1	1	1	18
菠菜	2	2	2	2	2	2	2	2	2	2	2	22
芹菜	2	1	1	1	1	2	1	2	2	2	1	16
蒜苗	2	2	2	2	2	1	1	1	2	2	2	19
茄子	2	1	1	1	1	2	1	1	1	1	1	13
黄瓜	1	1	1	1	1	1	1	1	1	1	1	11

综合考虑蔬菜品种的营养价值和常见度,定义蔬菜品种研究价值

$$RVal(i) = Com(i) \cdot NVal(i),$$

$RVal(i)$ 表示第 i 种蔬菜的研究价值, $Com(i)$ 表示第 i 种蔬菜的常见度系数,

$NVal(i)$ 表示第 i 种蔬菜营养价值。

代入数值, 计算结果如表 28 所示。

表 28 蔬菜评分计算结果

品种	$NVal$	Com	$RVal$
大白菜	18	0.1168	2.1027
萝卜	18	0.0450	0.8103
黄瓜	11	0.0488	0.5365
菠菜	22	0.0194	0.4257
蒜苗	19	0.0198	0.3753
茄子	13	0.0286	0.3713
芹菜	16	0.0230	0.3675
胡萝卜	18	0.0166	0.2990

按照研究价值从大到小排序, 挑选前 6 位的蔬菜品种作为最终的研究对象。即主要蔬菜为小白菜、大白菜、萝卜、黄瓜、菠菜、蒜苗、茄子。

4.4 估计主要蔬菜消费量

上文挑选了 6 种蔬菜, 它们的年产量数据见表 5-27

表 29 六中蔬菜年产量 (单位: 万吨)

year	萝卜	蒜苗	茄子	黄瓜	大白菜	菠菜
1990	1715	439	511	720	3555	281
1991	1287	497	579	758	3578	321
1992	1049	437	558	975	2673	293
1993	1015	539	653	1075	2916	330
1994	1216	494	759	1256	3034	324
1995	1512	612	910	1511	3830	463
1996	1809	730	1061	1766	4626	602
1997	2105	848	1212	2021	5422	741
1998	2401	966	1363	2276	6218	880
1999	2697	1084	1515	2531	7014	1018
2000	2993	1202	1666	2786	7810	1157
2001	3289	1320	1817	3041	8606	1296
2002	3585	1438	1968	3296	9401	1435
2003	3881	1556	2119	3551	10197	1574
2004	3833	1567	2191	3655	10346	1567
2005	3935	1654	2263	3817	10308	1618
2006	4003	1832	2247	4040	10506	1627
2007	3979	1830	2310	4130	10526	1645

2008	3955	1828	2372	4219	10546	1663
2009	4080	1790	2589	4420	10587	1754

4.4.1 多项式拟合法

在 Matlab 2010 编写程序对 6 种蔬菜的年产量及相应的年份进行多项式拟合，通过检测残差，合理选择拟合多项式的阶数。萝卜年产量的拟合结果如图 10 所示：

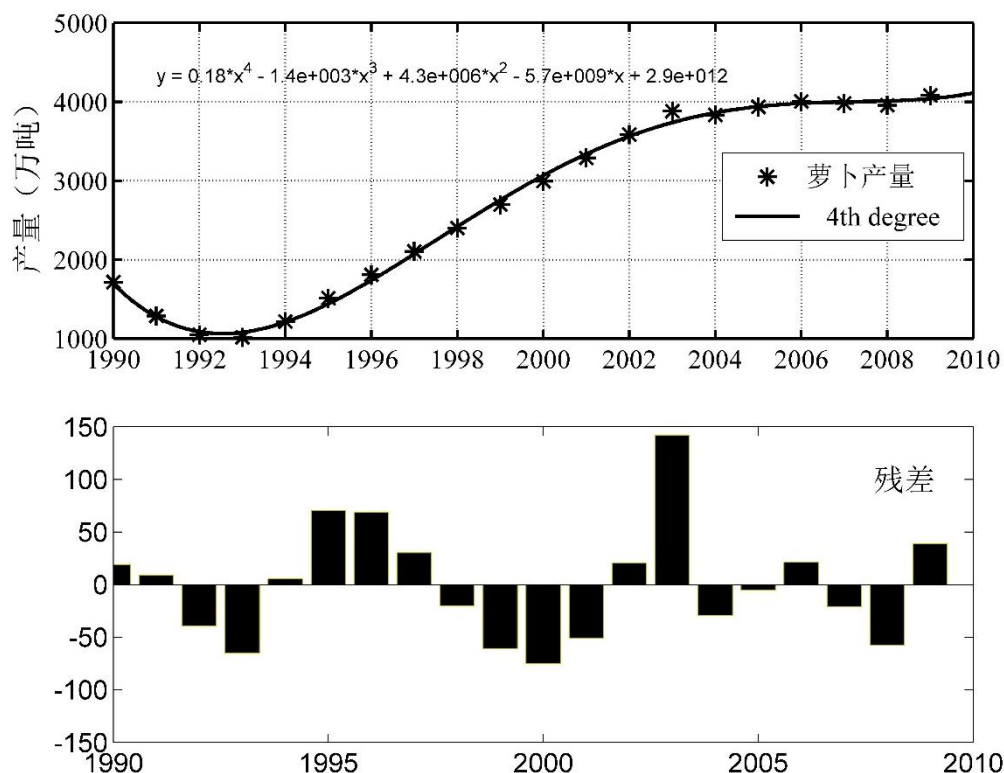


图 10 萝卜年产量拟合结果

采用 4 阶多项式对萝卜产量及年份数据进行拟合，从图中可以发现，在 (1990,2009) 年份区间内拟合曲线与原有数据点符合度很高，拟合效果非常好。

但是当超过原有数据范围后，拟合曲线的变化极其剧烈，以此来预测未来的产量存在明显误差，如图 11 所示，预测年份分别为 2015 年和 2020 年的产量情况，萝卜产量在 2010 年后迅速增大。在实际情况下，由于自然条件、国家政策等因素比较稳定，蔬菜的产量不可能无限制增长，预测得出的数据明显与前几年的数据走势不相符，所以采用多项式拟合的方法只能预测近几年的产量，不能作为长远预测的方法。

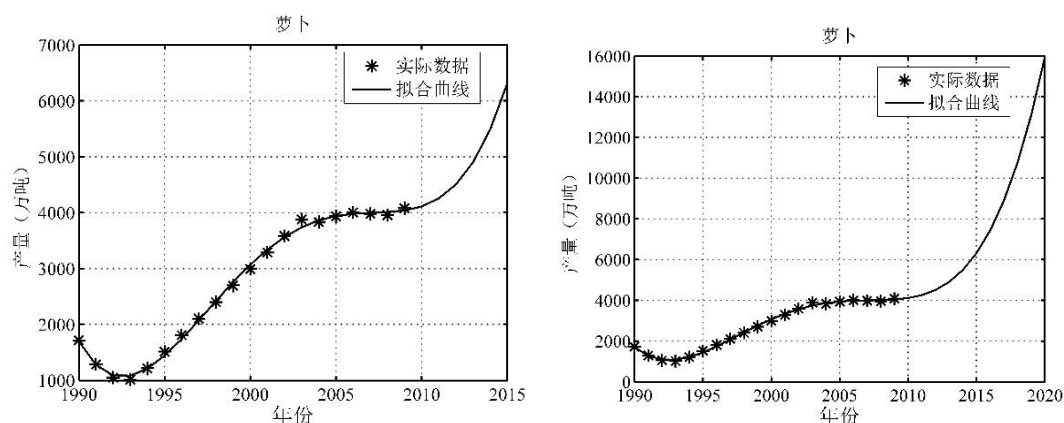


图 11 2015 年和 2020 年萝卜年产量

4.4.2 Logistic 复合模型法

蔬菜的 Logistic 复合模型法和水果的 Logistic 模型法计算过程一样，本小节对过程将不再重复想说。

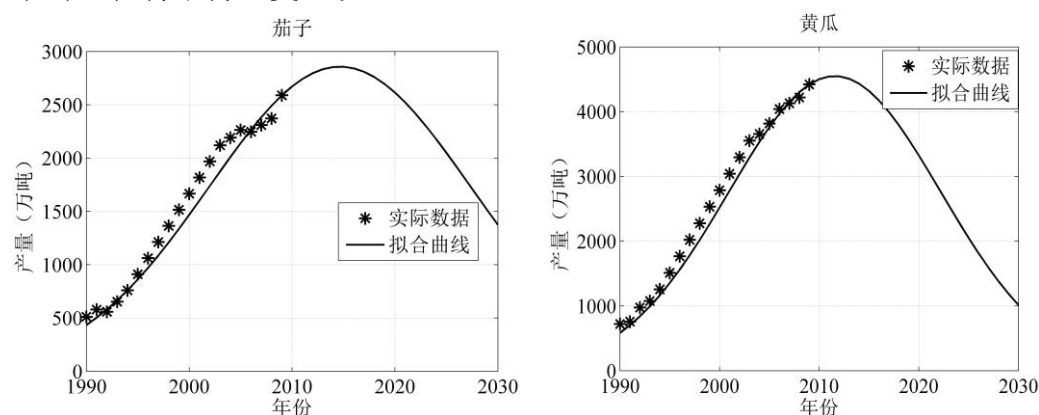


图 12 蔬菜 Logistic 模型预测产量

以此类推可以得到其他蔬菜的预测产量，对应地水果的消费量就可以得出，见表 30。

表 30 2005-2030 年蔬菜消费量

年份	萝卜	蒜苗	茄子	黄瓜	大白菜	菠菜
2005	2840	1267	1852	3431	5909	967
2006	3021	1336	1960	3620	6230	1031
2007	3218	1401	2062	3786	6555	1091
2008	3432	1462	2155	3924	6881	1145
2009	3666	1518	2239	4031	7208	1193
2010	3920	1567	2312	4105	7535	1233
2011	4198	1610	2372	4144	7859	1265
2012	4502	1646	2419	4146	8179	1287
2013	4835	1674	2451	4112	8494	1300
2014	5199	1693	2469	4042	8802	1302
2015	5598	1704	2471	3938	9101	1294
2016	6036	1706	2458	3803	9390	1277
2017	6517	1700	2430	3641	9668	1249

2018	7045	1684	2387	3455	9932	1213
2019	7628	1661	2330	3249	10181	1169
2020	8269	1629	2261	3029	10414	1118
2021	8977	1590	2180	2799	10629	1060
2022	9758	1543	2089	2563	10825	998
2023	10622	1490	1990	2327	11000	932
2024	11578	1432	1883	2094	11155	864
2025	12638	1369	1772	1868	11287	794

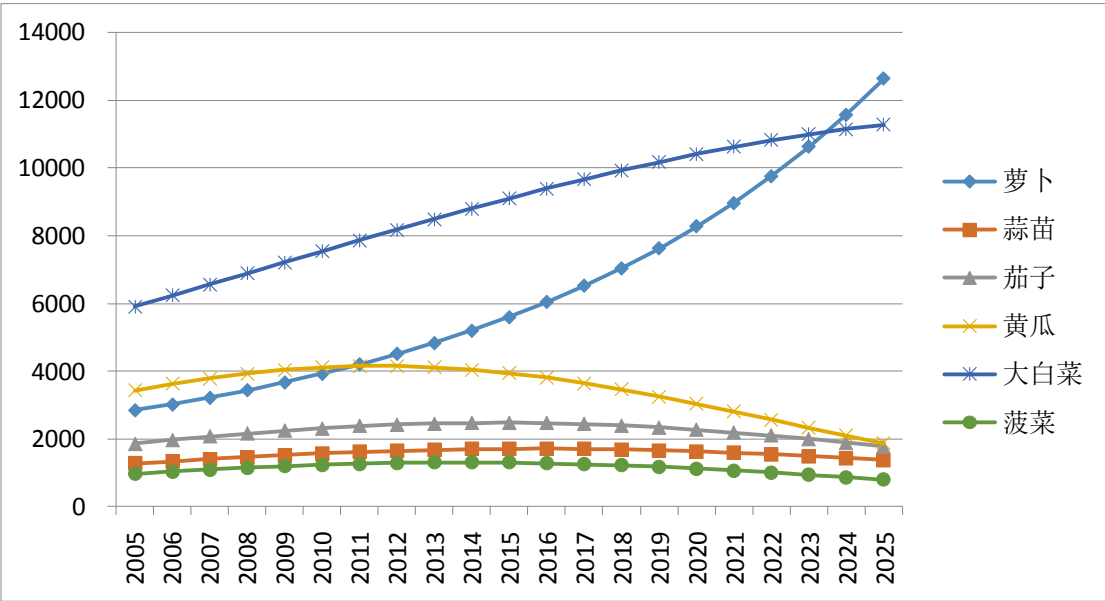


图 13 蔬菜消费量

用表 30 里的 2005-2025 年蔬菜消费量数据绘制得出图 13，由图 13 可以看出，从 2005 年到 2025 年，萝卜、大白菜的消费量一直呈现增加的趋势，并且萝卜的增长率大于大白菜；黄瓜先缓慢增长，到 2011 年达到峰值后开始减少；其余的茄子、蒜苗、菠菜刚开始缓慢增加后面缓慢减少，茄子、蒜苗、菠菜的消费量基本可认为处于稳定水平。

5 营养素摄入现状与摄入水平探究

5.1 营养素摄入现状

5.1.1 营养素摄入现状文献分析

人体需要的营养素主要有蛋白质、脂肪、维生素、矿物质、糖和水，其中维生素对于维持人体新陈代谢的生理功能是不可或缺的，另外膳食纤维对促进良好的消化和排泄固体有着举足轻重的作用。

因此，研究人体营养素的摄入现状就显得尤为重要，不少学者也对这一问题进行了广泛且深入的研究。何宇纳、杨桢等查采用多阶段整群抽样法，通过对北京市 6 个区县居民的营养素摄入情况调查得出结论：所调查的居民营养素摄入量未达到所推荐的范围；女性营养素的平均摄入量低于男性，但女性中营

养补充剂的使用率高于男性[4]。栗彦芳，李美英等选取广州市 141 名年龄在 20-50 岁，身体一切状况正常的居民作为研究对象，运用 24h 回顾调查法进行试验，最后得出结论：男性锌、硒、维生素 B1、维生素 B2 和维生素 C 摄入不足的比例分别为 35.8%、19.4%、71.6%、61.2%和 70.1%；女性分别为 6.8%、27.0%、59.5%、36.5%和 36.5%[5]。扈晓杰、韩冬、李铎在“膳食纤维的定义、分析方法和摄入现状”一文中指出，我国居民膳食纤维摄入量在 1989—1997 年间呈现下降趋势，成年男性居民平均每日膳食纤维摄入量从 17.0g/d 下降到 13.6 g/d,成年女性居民平均每日膳食纤维摄入量从 15.6g/d 下降到 11.7g/d。1997—2006 年居民平均每日膳食纤维摄入量基本稳定，男性约为 13.0g/d，女性约为 12.5g/d[6]。

由上述分析可知，不管是维生素、膳食纤维还是矿物质，我国居民都表现摄入量不够。为了更深入刻画居民具体的摄入量及摄入现状，需要采用问题一的水果蔬菜消费量的总体数据进行建模计算。

5.1.2 营养素摄入量建模分析

(1) 调整系数确定

中国居民营养素人均日摄入量标准考虑的是人从所有食物中摄入的营养素量，而从水果和蔬菜中摄入的营养成分只占其中的一小部分，所以要计算人从果蔬中摄入的营养素数量标准。用中国居民营养素人均摄入量乘以一定的调整系数 λ ，从而进行调整，进而得到人体从果蔬中日均摄入的营养素标准。我们需要明确的是有些营养素，例如钙，对于人体来说，营养的摄入主要源牛奶而非水果或蔬菜。为了确定调整系数 λ ，首先营养素的级别进行分级别。级别从 1 到 5，1 表示人体摄入这些营养大部分是通过食水果或蔬菜，5 表示这些营养素基本不靠水果或蔬菜提供。类别分出来后，不同类别我们给予不同数值的调整系数，见表 30,31。然后，利用调整系数计算人从果蔬中摄取的营养物标准，如表 32。

表 31 不同营养素的级别

营养素	级别
A	1
B1	3
B2	2
C	1
D	1
钙	2
铁	3
钠	4
锌	4
硒	5
膳食纤维	—

表 32 不同级别的调整系数

级别	调整系数
1	0.8
2	0.6

3	0.4
4	0.2
5	0.1

结合上面两个表格，可以综合得出表 33。

表 33 人从果蔬中摄入营养素标准

营养物质	级别	调整系数	人体需求标准 (平均)	从果蔬中摄取 标准
A	1	0.8	764.48	611.5808
B1	3	0.4	1.30	0.52016
B2	2	0.6	1.33	0.80016
C	1	0.8	102.47	81.976
D	1	0.8	12.64	10.11104
钙	2	0.6	882.34	529.404
铁	3	0.4	14.96	5.98284
钠	4	0.2	1983.04	396.6076
锌	4	0.2	15.40	3.08082
硒	5	0.1	46.45	4.64476
膳食纤维	—	—	>0	>0

(2) 居民营养素摄入量计算

设第 i 种水果的总消费量为 $Con_1(i)$ 万吨，第 j 种蔬菜的总消费量为 $Con_2(j)$ 万吨，第 i 种水果所含第 j 种营养素的质量是 $N_1(i, j)$ 微克，第 i 种蔬菜所含第 j 种营养素的质量是 $N_2(i, j)$ 微克，则按照第一问中筛选出来的主要水果和蔬菜来计算中国居民目前第 j 种营养物质的年摄入量 γ 的数学表达式为（单位为 ug）：

$$\gamma = \sum_{i=1}^8 \frac{Con_1(i) \cdot 10^{10}}{13 \times 10^8 \times 365} \cdot \frac{N_1(i, j)}{10^2} + \sum_{i=1}^6 \frac{Con_2(i) \cdot 10^{10}}{13 \times 10^8 \times 365} \cdot \frac{N_2(i, j)}{10^2} \quad j=1, 2, \dots, m. \quad (14)$$

5.2 我国居民营养素摄入水平分析

5.2.1 我国居民营养素摄入量计算

按照上一节给出的公式（14），代入各种水果和蔬菜的消费量，从而可以计算出每种营养素在各年份的摄入量。为方便分析，现列出 2010-2014 年各营养素的摄入量 γ ，见表 34。

表 34 2010-2014 年居民从果蔬中摄入营养素量

年份	A	B1	B2	C	E	钙	铁	钠	锌	硒	膳食纤维
2010	457.45	0.25	0.33	170.47	5.83	248.44	3.52	224.39	1.78	3.67	3.55

2011	481.05	0.26	0.34	180.23	6.09	260.16	3.65	234.90	1.88	3.85	3.68
2012	503.56	0.27	0.36	189.95	6.34	271.74	3.77	245.45	1.98	4.02	3.79
2013	524.76	0.28	0.37	199.55	6.58	283.16	3.88	256.01	2.07	4.18	3.89
2014	544.48	0.29	0.38	208.98	6.81	294.43	3.98	266.59	2.16	4.34	3.98

5.2.2 达标指数确定

我们在处理保障人体健康所需要的各种营养成分的范围时，充分考虑到各个年龄段对个营养成分的需求是不同的，并且考虑到个年龄段的人数比例来确定最终的营养范围标准，具体做法是取各年龄层的加权平均值。首先我们根据题目附件的，找到不同年龄段对营养素的需求量。然后加权得到人均不同营养素的应摄入量 θ 。数据如表 35，结果如表 36。

表 35 不同人群平均营养素标准

年龄段	人数 比例 (%)	A	B1	B2	C	E	钙	铁	钠	锌	硒
0-14	16.6	586	0.8	1.0	70	5.8	740	11.6	893	11.6	28.6
15-64	74.5	800	1.4	1.4	110	14	900	15.7	2200	16.3	50

表 36 人体营养素应摄入量

年龄段	人数 比例 (%)	A	B1	B2	C	E	钙	铁	钠	锌	硒
平均值		764.5	1.3	1.3	102.5	12.6	882.3	15.0	1983	15.4	46.4

为评价摄入量水平，我们引入达标指数 β 这一指标，表示人实际从果蔬中摄入的营养素的量与人应该从果蔬中摄入量的比值，即：

$$\beta = \frac{\gamma}{\lambda \theta} \quad (15)$$

5.2.3 我国居民摄入营养素水平分析

结合 λ 取值表 31、 μ 计算数值表 33 和人体不同营养素应摄入量 θ 表 36，可以计算得出 2010、2011、2012、2013、2014、2020 年不同营养素的达标指数，见表 37。

表 37 不同年份的营养素达标指数表

营养素	2010	2011	2012	2013	2014	2020
A	0.74798	0.786568	0.823374	0.858039	0.890283	1.02446
B1	0.480621	0.499846	0.519071	0.538296	0.557521	0.67287
B2	0.412418	0.424915	0.44991	0.462408	0.474905	0.574885
C	2.079511	2.19857	2.317142	2.434249	2.549283	3.153118
D	0.576597	0.602312	0.627037	0.650774	0.673521	0.786269
钙	0.469282	0.491421	0.513294	0.534866	0.556154	0.683051
铁	0.588349	0.610078	0.630136	0.648521	0.665236	0.718722

钠	0.565773	0.592273	0.618874	0.645499	0.672176	0.836948
锌	0.577768	0.610227	0.642686	0.671899	0.701112	0.840685
硒	0.790138	0.828891	0.865491	0.899939	0.934386	1.123847

计算出达标指数后，就可以运用达标指数与 1 进行对比；若小于 1，则表示摄入量不够；若大于 1，则可认为是摄入量富足。并根据达标指数数值，画出达标指数的折线图，见图 13。

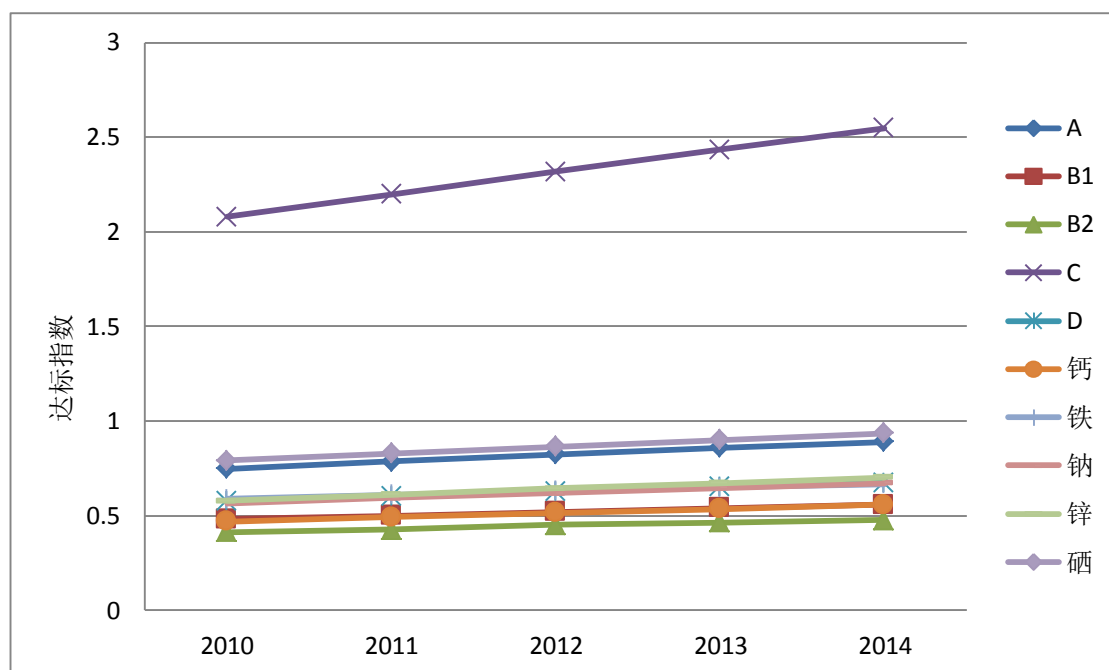


图 13 2010-2014 达标指数图

由上图可知，除了维生素 C，其余的 β 均小于 1。这表明我国居民目前维生素 A、B1、B2、D、钙、铁、钠、锌、硒的摄入量都低于应该摄入量，摄入水平表现为不合理。维生素 C 摄入量高于应该摄入量，摄入水平合理。

5.3 2020 年我国居民营养健康状况

由表 36 可知，

(1) 2020 年，有维生素 A 和有三种营养素的摄入达标指数大于 1，分别为维生素 A、维生素 C 和硒。而对于 2014 年，仅仅只有维生素 C 达到应摄入量。

(2) 从 2010 到 2014 年，摄入达标指数一直增加，且 2020 年各个营养素的摄入达标指数都比 2010-2014 年大。

(3) 到 2020 年还是有很多营养素不满足该摄入量，表明到 2020 年居民的营养素情况还不合理。

由此可以看出，到 2020 年，我国居民的营养素摄入量水平相比现在而且趋于好，但也没有达到标准。

6 居民果蔬人均消费优化方案

对于本问题，首先，我们将全国分为南北两个区域，并按照春、夏、秋、冬四季来研究该问题。为综合考虑这个问题，我们引入应地系数和应季系数来衡量地区和季节因素对价格的影响。然后，利用 Matlab 求解该优化模型的解，给出南北方四季的主要水果和蔬菜的按年度合理的人均消费量。最后，利用聚类分析方法，对不同的水果和蔬菜按照营养素种类进行聚类，进而对优化模型得到的果蔬方案进行各种组合搭配，并给出按照价格从小到大的替代方案。这样，居民可以按照各种喜好搭配自己的饮食方案，又可以选择到价格较低的搭配方案。本问题的技术路线图如图 14。

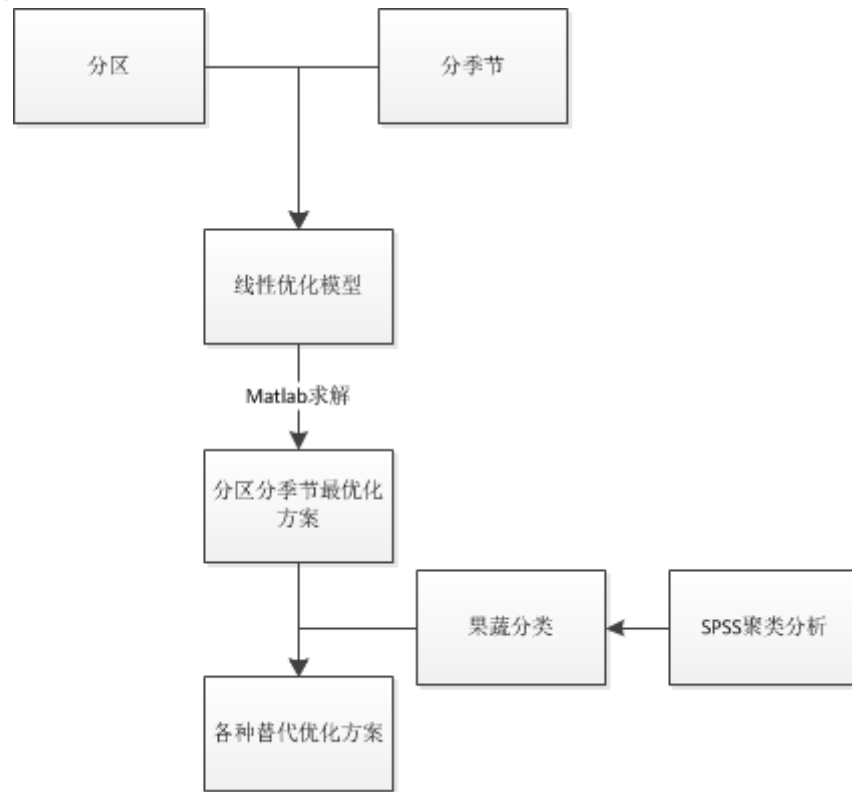


图 14 问题三的技术路线图

对于现如今我国居民来说，在保证营养均衡满足健康需要条件下，提供主要的水果和蔬菜产品的按年度合理人均消费量，使人们能够以较低的购买成本，我们考虑这是一个线性优化问题。

该模型的表达式为：

$$\begin{aligned}
 & \text{Min} \quad \sum_{i=1}^{14} p_i x_i \\
 & \text{s.t.} \quad \sum_{i=1}^{14} a(i, j) x_i \geq b_j \quad j=1, 2, \dots, 10 \\
 & \quad \quad \sum_{i=1}^{14} a(i, 11) x_i > 0 \\
 & \quad \quad 0 \leq x_i \leq 0.5 \quad i=1, 2, \dots, 14
 \end{aligned} \tag{16}$$

其中 p_i 代表第 i 种果蔬每千克的价格， $a(i, j)$ 代表第 i 种果蔬第 j 种营养成分的含量， $i = 1, 2, \dots, 14; j = 1, 2, \dots, 11$ 。

6.1 居民果蔬人均消费优化准备

6.1.1 分区域分季节

首先，综合考虑全国范围内各地的果蔬价格（价格行情网），确定基准价格，根据果蔬种植区域和价格特点等，我们把全国分为南方和北方两个区域，季节按照春季、夏季、秋季、冬季来考虑。各种水果和蔬菜的基准价格、种植区域以及所属季节如表 38 所示：

表 38 种植地域和所属季节

品种	基准价格（元/kg）	种植地域	所属季节
苹果	9.0	北	秋
橘子	6.3	南	秋
梨	9.0	北	秋
香蕉	12.8	南	夏
桃子	5.6	南	夏
葡萄	11.3	北	夏
大枣	13.0	北	秋
柿子	9.5	北	秋
萝卜	2.0	南	冬
蒜苗	12.5	北	夏
茄子	4.1	南	夏
黄瓜	5.0	南	夏
大白菜	2.0	北	秋
菠菜	11.0	南	春

6.1.2 应地系数和应季系数的引入

考虑到水果和蔬菜的价格会受到地区和季节的双重影响，引入应地系数和应季系数来调控不同地区不同季节的价格以反映当地当时的果蔬价格变化。

应地系数的含义是种植在当地的果蔬的价格会偏低，应地系数小于 1，不种植在当地的果蔬价格会偏高，应地系数大于 1；应季系数的含义是在当季生产的果蔬的价格会偏低，应季系数小于 1，不在当季生产的果蔬价格会偏高，应地系数大于 1。系数取值见表 39。

表 39 系数取值

系数	应地	不应地	应季	不应季
取值	0.8	1.2	0.8	1.2

利用应地系数和应季系数，得到居民在不同地区不同季节购买不同果蔬的价格变化，例如，南春代表在南方春季来购买该种果蔬的价格。具体价格见表 40。

表 40 不同购买情况的价格

品种	基准价格	南春	南夏	南秋	南冬	北春	北夏	北秋	北冬
苹果	9.00	12.96	12.96	8.64	12.96	8.64	8.64	5.76	8.64
橘子	6.30	6.05	6.05	4.03	6.05	9.07	9.07	6.05	9.07
梨	9.00	12.96	12.96	8.64	12.96	8.64	8.64	5.76	8.64
香蕉	12.80	12.29	8.19	12.29	12.29	18.43	12.29	18.43	18.43
桃子	5.60	5.38	3.58	5.38	5.38	8.06	5.38	8.06	8.06
葡萄	11.30	16.27	10.85	16.27	16.27	10.85	7.23	10.85	10.85
大枣	13.00	18.72	18.72	12.48	18.72	12.48	12.48	8.32	12.48
柿子	9.50	13.68	13.68	9.12	13.68	9.12	9.12	6.08	9.12
萝卜	2.00	1.92	1.92	1.92	1.28	2.88	2.88	2.88	1.92
蒜苗	12.50	18.00	12.00	18.00	18.00	12.00	8.00	12.00	12.00
茄子	4.10	3.94	2.62	3.94	3.94	5.90	3.94	5.90	5.90
黄瓜	5.00	4.80	3.20	4.80	4.80	7.20	4.80	7.20	7.20
大白菜	2.00	2.88	2.88	1.92	2.88	1.92	1.92	1.28	1.92
菠菜	11.00	7.04	10.56	10.56	10.56	10.56	15.84	15.84	15.84

6.2 居民果蔬人均消费优化求解

本小节对线性规划模型（16）利用 Matlab 来计算该优化模型的结果如下：

（1）在南方地区建议果蔬的按年度合理的人均消费如表 41。

表 41 在南方买水果优化结果

季节	购买的果蔬的种类	消费量（Kg）	总价（元）
春季	橘子	14.6	1281.15
	萝卜	182.5	
	大白菜	182.5	
	菠菜	43.8	
夏季	橘子	18.25	1357.8
	萝卜	182.5	
	茄子	73	
	大白菜	182.5	
	菠菜	18.25	
秋季	橘子	14.6	1230.05
	萝卜	182.5	
	大白菜	182.5	
	菠菜	43.8	
冬季	橘子	14.6	1317.65
	萝卜	182.5	
	茄子	14.6	
	大白菜	182.5	
	菠菜	40.15	

（2）在北方地区建议果蔬的按年度合理的人均消费如表 42：

表 42 在南方买水果优化结果

季节	购买的果蔬的种类	消费量 (Kg)	总价 (元)
春季	大枣	7.3	1463.65
	萝卜	182.5	
	蒜苗	3.65	
	大白菜	182.5	
	菠菜	43.8	
夏季	橘子	21.9	1595.05
	萝卜	182.5	
	蒜苗	40.15	
	大白菜	182.5	
	菠菜	14.6	
秋季	橘子	10.95	1551.25
	萝卜	182.5	
	大白菜	182.5	
	菠菜	43.8	
冬季	大枣	7.3	1522.05
	萝卜	182.5	
	蒜苗	3.65	
	大白菜	182.5	
	菠菜	43.8	

6.3 对水果和蔬菜进行聚类分析

人们往往会碰到通过划分同种属性的对象很好地解决问题的情形，而不论这些对象是个体、公司、产品甚至行为。如果没有一种客观的方法，基于在总体内区分群体的战略选择，比如市场细分将不可能，其他领域也会遇到类似的问题，研究者都在基于一个多维剖面的观测中寻找某种“自然”结构。为此最常用的技巧是聚类分析，聚类分析将个体或对象分类，使得同一类中的对象之间的相似性比与其他类的对象的相似性更强。目的在于使类间对象的内同质性最大化和类与类间对象的异质性最大化。

6.3.1 欧氏距离

在多指标统计分析中,距离的概念十分重要,样品间的不少特征都可用距离去描述。大部分多元方法是建立在简单的距离概念基础上的。即平时人们熟悉的欧氏距离，或称直线距离。

就大部分统计问题而言，欧氏距离是不能令人满意的。这里因为，每个坐标对欧氏距离的贡献是同等的。当坐标轴表示测量值时，它们往往带有大小不等的随机波动，在这种情况下，合理的办法是对坐标加权，使得变化较大的坐标比变化小的坐标有较小的权系数，这就产生了各种距离。欧氏距离还有一个缺点，这就是当各个分量为不同性质的量时，“距离”的大小竟然与指标的单位有关。

6.3.2 马氏距离

最常用的一种统计距离是印度统计学家马哈拉诺比斯（*Mahalanobis*）于 1936 年引入的距离，称为“马氏距离”。设 X 、 Y 从均值向量为 μ ，协方差阵为 Σ 的总体 G 中抽取的两个样品，定义 X 、 Y 两点之间的马氏距离为：

$$d_m^2(X,Y)=(X-Y)\sum^{-1}(X-Y)$$

X 与总体 G 的马氏距离为：

$$d_m^2(X,G)= (X-\mu)\sum^{-1}(X-\mu)$$

马氏距离的最大优点是距离量纲无关,可以排除变量之间的相关性的干扰。因此本节中的聚类方法使用的就是马氏距离。

6.3.3 运用 SPSS 对果蔬聚类

利用 SPSS 对水果和蔬菜进行聚类分析，将食用功能大致相近的水果蔬菜分成一类，最后聚类结果为下图 15 所示，其中 1-14 是指代苹果、橘子、香蕉、桃子、葡萄、大枣、柿子、萝卜、蒜苗、黄瓜、大白菜。

树状图

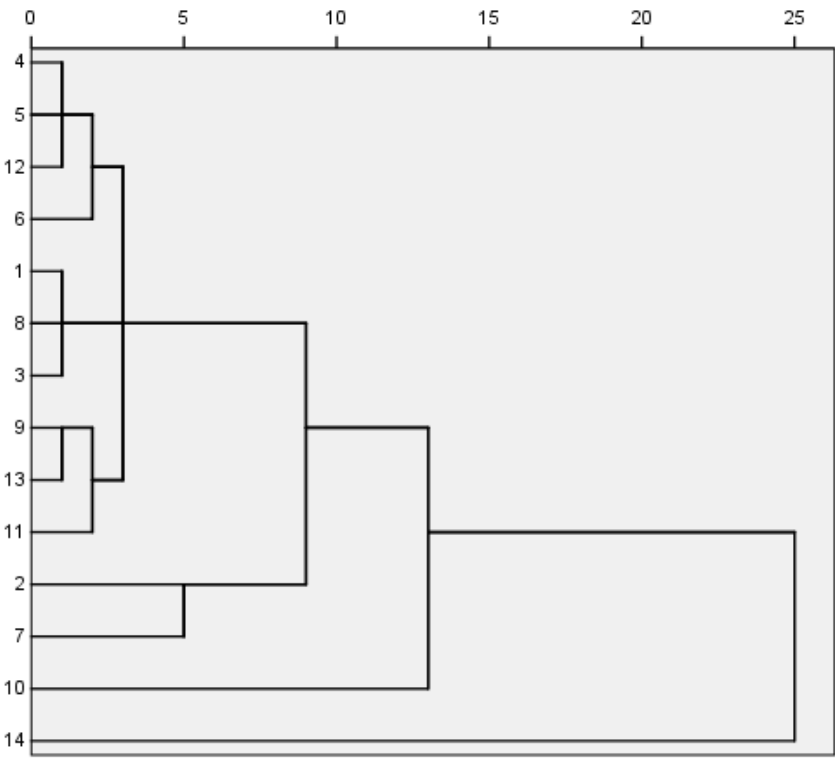


图 15 spss 聚类结构图

表 43 显示的是聚类的分类结果：

表 43 聚类结果

类别数	分类结果
1	香蕉、桃子、黄瓜、葡萄、苹果、柿子、梨、萝卜、大白菜、茄子

2	橘子、大枣
3	蒜苗
4	菠菜

6.4 我国居民人均果蔬消费搭配与替换方案

利用上述结果，可以对不同地区在不同季节的水果蔬菜食用品种进行替换和搭配，具体结果如下：

表 44 南方可替代或搭配方案

季节	购买的果蔬的种类	替代物（按照价格从小到大排序）
春季	橘子	大枣
	萝卜	茄子、黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子
	大白菜	茄子、黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子
	菠菜	—
夏季	橘子	大枣
	萝卜	黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子
	茄子	黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子
	大白菜	黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子
	菠菜	—
秋季	橘子	大枣
	萝卜	茄子、黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子
	大白菜	茄子、黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子
	菠菜	—
冬季	橘子	大枣
	萝卜	黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子
	茄子	黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子
	大白菜	黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子
	菠菜	—

表 45 北方可替代或搭配方案

季节	购买的果蔬的种类	替代物（按照价格从小到大排序）
春季	大枣	橘子
	萝卜	茄子、黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子
	蒜苗	—
	大白菜	茄子、黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子
	菠菜	—
夏季	橘子	大枣
	萝卜	黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子
	茄子	黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子
	大白菜	黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子
	菠菜	—
秋季	橘子	大枣
	萝卜	茄子、黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子
	大白菜	茄子、黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子

	菠菜	—
冬季	橘子	大枣
	萝卜	黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子
	茄子	黄瓜、桃子、香蕉、葡萄、苹果、梨、柿子
	大白菜	—
	菠菜	—

7 问题四模型建立与求解

问题四是一个多目标规划问题，需要在综合考虑居民人体营养均衡的情况下，一方面使得居民购买成本达到最低，另一方面使得种植者的利润最大，可以通过线性加权法求解。为了获得国家对于水果和蔬菜各品种的生产规模的战略调整方案，对 2015-2020 年的果蔬产品的年度合理人均消费量进行求解，考虑营养均衡、土地面积等因素的影响，可以认为在国家对果蔬各品种调控前后的年总消费量不变，仅仅改变了各种果蔬品种的比例。为了达到居民人体营养均衡的目的，可以将各项营养素的达标指数作为动态约束条件，使得居民各项营养素摄入量可以逐步均衡，最终实现全部达标的目的。

7.1 多目标规划模型的建立

为了对水果和蔬菜各品种的生产规模做出战略性调整，除考虑居民购买成本、种植者收益最优化外，还要考虑居民人体的营养均衡、土地面积等其他因素，只有对各种因素的指标进行综合衡量后，才能做出合理的决策，这种问题是一种多目标最优化问题。针对该问题的数学模型为

$$\begin{aligned}
 & \min \quad \sum_{i=1}^{14} p_i x_i \\
 & \max \quad \sum_{i=1}^{14} p_i r_i x_i \\
 & \text{s.t.} \quad \beta_{\min j} b_j \leq \sum_{i=1}^{14} a(i, j) x_i \leq \beta_{\max j} b_j \quad j=1, 2, \dots, 10 \\
 & \quad \sum_{i=1}^{14} a(i, 11) x_i > 0 \\
 & \quad \sum_{i=1}^{14} x_i \leq \sum_{k=1}^{14} con_k
 \end{aligned} \quad (17)$$

其中决策变量为 14 种果蔬的年度人均消费量 x_i 千克， $i=1, 2, \dots, 14$ ，两个目标函数分别为：

- (1) 居民购买果蔬的成本最小，其中 p_i 为对应的果蔬价格；
- (2) 种植者所获得的利润最大，其中 r_i 为对应果蔬的利润率。

约束条件分别为：

(1) 营养素摄入尽量均衡，对果蔬的品种进行调整，使通过国家政策宏观调控果蔬品种，得 2015-2020 年各种营养素达标指数逐步提升。

$$\beta_{\min j} \leq \beta_j \leq \beta_{\max j}, j = 1, 2, \dots, 10$$

$$\text{其中, } \beta_j = \frac{\sum_{i=1}^{14} x_i \cdot a(i, j)}{b_j}, j = 1, 2, \dots, 10 \text{ 为营养素达标指数。}$$

根据 2014 年的各项营养素达标指数，不断调整 $\beta_{\min j}$ 、 $\beta_{\max j}$ 大小，使得所有的营养素达标指数能够在 2020 年都超过 1，即实现营养素均衡。

(2) 考虑所有果蔬年度消费总量不变，年度消费量和土地面积成正比例关系，因此此约束相当于考虑土地面积因素的限制。

求解多目标规划问题的方法很多，主要有理想点法、线性加权法、最大最小法等。这里主要采用线性加权法，并且将每一年的权重看成随着时间动态变化的函数。

7.2 动态线性加权法及其实现

在具有多个指标的问题中，我们总希望那些相对重要的指标给与较大的权系数，因而将多目标问题转化为所有目标的加权求和的标量问题。基于这个现实，构造评价函数，即

$$\min_{x \in \Omega} f(x) = \sum_{i=1}^m \omega_i F_i(x)^2$$

其中 ω_i 为加权因子，其选取的方法很多，有专家打分法、容限法和加权因子分解法等。该模型的加权因子考虑了国家宏观调控政策，兼顾了消费者和生产者双方的利益，该因子并随着时间动态变化，如下表所示。在初期优先考虑消费者的利益，以扩大消费市场，所以初始时 $\omega_1 > \omega_2$ ，然后逐步平衡消费者和种植者的利益，所以到 2020 年时， $\omega_1 = \omega_2 = 0.5$ 。

表 46 ω_1 , ω_2 的取值

年份	ω_1	ω_2
2015	0.60	0.40
2016	0.58	0.42
2017	0.56	0.44
2018	0.54	0.46
2019	0.52	0.48

2020	0.50	0.50
------	------	------

7.3 动态多目标规划模型的求解

首先，求解 2015 年各果蔬品种年度合理人均消费量，根据 2014 年各营养素的达标指数，设定优化模型中约束条件， $\beta_{\min j}$ 、 $\beta_{\max j}$ 。由前文可知，在 2014 年，维生素 C 的达标指数为 2.55，说明在当前的果蔬品种消费配比情况下，居民维生素 C 的摄入量明显过剩，我们可以通过减小维生素 C 的达标指数，来控制果蔬品种配比，使得其他未达标的营养素能够在总消费量逐年增长的情况下更快的达标。通过反复试验，设定

$$\beta_{\min j} = [0.89 \quad 0.56 \quad 0.48 \quad 1 \quad 0.67 \quad 0.56 \quad 0.66 \quad 0.67 \quad 0.7 \quad 0.94];$$

$$\beta_{\max j} = [\infty \quad \infty \quad \infty \quad 2.3 \quad \infty \quad \infty \quad \infty \quad \infty \quad \infty \quad \infty];$$

将维生素 C 的达标指数限定在 2.3 以下。

通过 Matlab 中的线性规划函数 linprog 求解出 x_i 。

通过公式 $\beta_j = \frac{\sum_{i=1}^{14} x_i \cdot a(i, j)}{b_j}$, $j=1,2,\dots,10$ ，得出 2015 年的各营养素达标指数

$$\beta_j = [0.89 \quad 0.48 \quad 0.61 \quad 2.3 \quad 0.74 \quad 0.82 \quad 0.66 \quad 1.17 \quad 0.7 \quad 0.94]$$

与 $\beta_{\min j}$ 的各项数值进行比较，如果数值已经超过 1，则说明该项营养素已经达标，不需要进一步提升，则在计算下一年的年度人均消费量时，可以设定对应的 $\beta_{\min j}$ 为 1。其他的还未达标的各项可以通过实验法，在以得出的 β_j 基础上加上一个较小的值，保证整个线性规划有解存在，即可行域不为空。

计算 2016 年的情况时，根据以上原则，设定

$$\beta_{\min j} = [0.89 \quad 0.6 \quad 0.6 \quad 1 \quad 0.76 \quad 0.8 \quad 0.7 \quad 0.67 \quad 0.75 \quad 0.95]$$

$$\beta_{\max j} = [\infty \quad \infty \quad \infty \quad 2.1 \quad \infty \quad \infty \quad \infty \quad \infty \quad \infty \quad \infty]$$

计算得出 2016 年最优解，并得出 2016 年的各营养素达标指数

$$\beta_j = [1.14 \quad 0.6 \quad 0.68 \quad 2.1 \quad 0.85 \quad 0.8 \quad 1.08 \quad 1.12 \quad 0.75 \quad 1.22]$$

重复采用上述方法依次计算 2017-2020 年的情况，确保各项营养素达标指数不断增加，尽量使得每一项都趋近于 1，得到 2015-2020 年的优化结果，见表 46。

表 47 2015-2020 年的优化结果

年份	人均居民最小成本	人均种植者最大利润
----	----------	-----------

2015	873	298
2016	1264	397
2017	1306	400
2018	1345	418
2019	1525	461
2020	2711	775

表 48 最优解

果蔬	2015	2016	2017	2018	2019	2020
苹果	0	0	0	0	0	0
橘子	27	13	4	0	0	0
梨	0	0	0	0	0	0
香蕉	0	0	0	0	0	0
桃子	0	0	0	0	0	0
葡萄	0	0	0	0	0	0
大枣	0	0	0	5	0	0
柿子	0	0	0	0	0	0
萝卜	184	191	132	61	16	0
蒜苗	2	33	28	5	0	62
茄子	0	0	0	0	0	0
黄瓜	0	0	0	0	0	0
大白菜	46	18	86	161	194	97
菠菜	20	32	45	71	101	159

四种主要的果蔬年度合理人居消费量如图 16 所示。

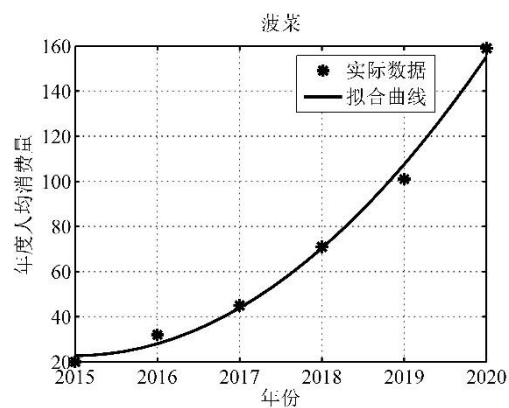
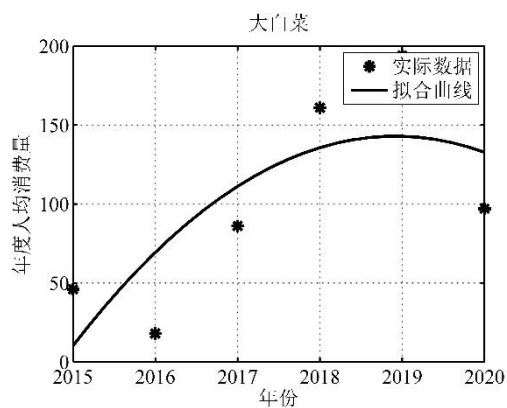
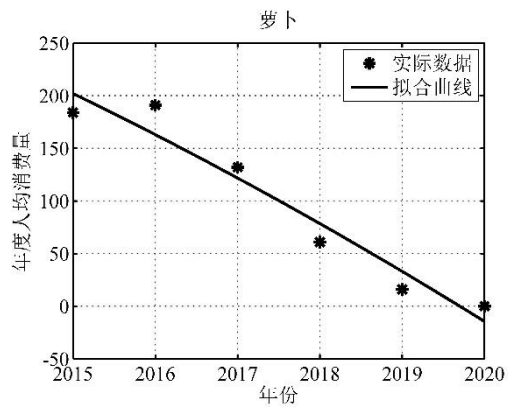
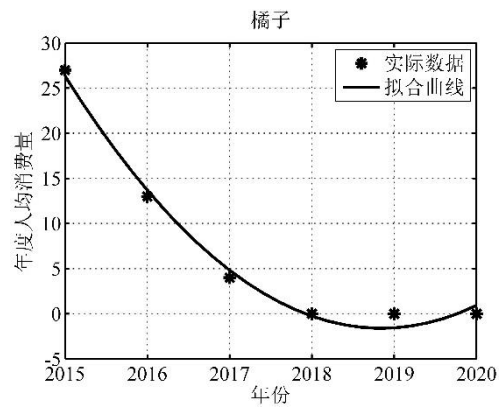


图 16 四种主要的果蔬年度合理人居消费量

7.4 国家宏观调控战略

通过问题一的求解可以发现，从 2015-2020 年，水果和蔬菜各个品种的产量呈现逐渐上升趋势，但是由于土地面积、自然环境等多方面的因素的影响，在未来一定时期，产量一定会达到饱和，在按照预测，到 2020 年中国居民各项营养素摄入量仍有多项没有达标。我们假设每一年果蔬总产量都是确定的，国家进行宏观调控可以改变各种果蔬产品的配比，来改善居民从果蔬摄入的营养平衡，同时兼顾居民的购买成本和种植者的收益。根据本题中线性加权法求解的结果，可以得出四种果蔬消费趋势如图 17 所示。我国的宏观调控应该顺应消费趋势，所以未来几年应该逐步减少橘子和萝卜的种植，加大菠菜和大白菜的种植面积。

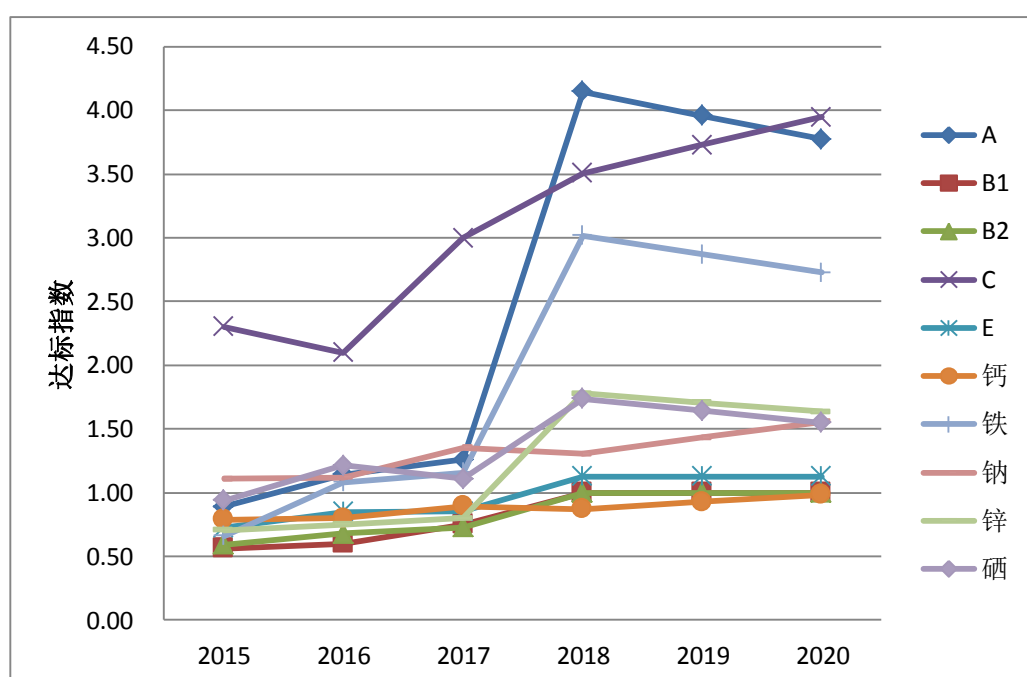


图 17 四种果蔬消费趋势

8 政策建议

8.1 政策建议

根据问题二得出的结论，目前我国居民大部分营养素摄入量严重不足，但是也存在部分营养素摄入过量的情况，说明我国居民在平衡膳食营养素方面不够重视，需要加强对居民食物与营养的指导，提倡健康生活方式。

根据问题四得出的结论，按照目前的果蔬发展现状，果蔬品种并不是十分合理，需要国家对我国的果蔬品种结构进行宏观调控。通过国家政策，逐步调整果蔬品种的比例，可以有效地改善目前我国居民营养素摄入不均的情况，也可以让中国居民以更加低廉的价格购买营养素含量更加均衡的果蔬品种，同时

兼顾果蔬种植者的利益，给予他们一些战略上的帮助和政策上的鼓励，帮助他们合理的选择果蔬品种。

问题四的结论表明，国家完全可以通过以往的数据预估未来一段时间内的果蔬消费水平，只要国家对果蔬的生产规模进行合理干预，可以平衡中国居民各种营养素的摄入水平，那么到 2020 年居民营养摄入量都能满足基本要求。从而可以看出在果蔬的调节中，国家及有关部门干预的积极作用。有关部门需要从以下几个方面开展工作：

(1) 有效把握果蔬产量等数据，做好分析预测，进而调整果蔬生产战略。

从历史数据中可以大致把握果蔬产品的发展状况，但是选取数据的不完善或者不准确可能导致对未来预测的不准确。因此，做好果蔬产量等数据记录和收集工作尤为重要，是一切战略政策制定的基础。

(2) 全面普及膳食营养和健康知识。

要加强对居民食物和营养知识的指导，提高全民营养意识，提倡健康生活方式，树立科学饮食理念。进而可以影响到人民的饮食习惯，避免再出现厌食、费食。发挥主要媒体对食物与营养知识进行公益宣传的主渠道作用，增强营养知识传播的科学性。加大对食物与营养事业发展的投入，加强流通、餐饮服务等基础设施建设。

(3) 对果蔬的生产与供给加以调控。

对白菜、萝卜类的营养素丰富且价格低的果蔬品种给予一定的政策鼓励，使得种植者扩大其生产量，满足更多居民的营养所需。

(4) 加大营养素含量监测与干预。

开展全国居民营养与基本健康监测工作，进行食物消费调查，定期发布中国居民食物消费与营养健康状况报告，引导居民改善食物与营养状况。

(5) 推进食物与营养法制化管理。

在食物与营养的工作上，提出并抓紧立法准备，保障食品安全，营造安全、诚信、公平的市场环境。在食品安全工作上要加大力度，使国民放心地购买果蔬，从而增加果蔬的消费量。

(6) 调整果蔬的进出口贸易量和价格

在我国果蔬的进出口贸易发展中，进口水果总量虽然与出口总量几乎持平，但是进口果蔬的价格要远远高于出口果蔬价格，这在一定程度上影响了国内的果蔬行业的发展，并使果蔬种植者的利益受到影响。所以，相关部门应一定程度上限制进口的果蔬量，并控制其价格。

9 模型评价

9.1 模型的优点

(1) 利用 Logistic 模型来预测果蔬产量，充分考虑到外界因素对产量的阻滞作用，可以长期预测果蔬产量的变化趋势。

(2) 问题一中筛选果蔬时，不仅充分考虑到营养素的成分和含量因素，并对其进行打分和分级，还利用果蔬的“常见度”进一步筛选，既定量衡量，又符合实际情况。

(3) 在问题三中，引入应季系数和应地系数，充分考虑到区域和季节对果

蔬价格的影响，使得结果更具现实意义。

（4）问题四的多目标优化问题中，将营养素达标指数作为动态约束条件，充分考虑了国家宏观调控的作用，并兼顾了消费者和生产者双方的利益。

9.2 模型的缺点

（1）问题一中的回归模型只能用于短期预测，无法对长期的果蔬产量进行预测。

（2）问题四中的多目标优化模型没有考虑进出口贸易的影响作用。

参考文献

- [1] 向云, 陆倩, 祁春节, 基于灰色马尔科夫模型对中国的水果产量预测研究, 农业生产展望, 4:43-44, 2014。
- [2] 淘建平, 熊刚初, 我国水果消费水平与城市化的相关性分析, 中国农村经济, 6:18-24, 2004。
- [3] 张峭, 杨霞, 中国水果消费现状分析及其预测, 产品预测, 8:18-19, 2004。
- [4] 何宇纳, 杨桢, 北京市成年人膳食及营养补充剂中维生素和矿物质的摄入状况, 营养学报, 30(2):125-125, 2008。
- [5] 栗彦芳, 李美英, 成年人膳食矿物质摄入与机体矿物质营养状况的关系初探, *Acta Nutrimenta Sinica*, 3(34), 2012。
- [6] 扈晓杰, 韩冬, 李铎, 膳食纤维的定义、分析方法和摄入现状, 中国食品学报, 3(11), 2011。